

Explicitly searching for useful inventions: dynamic relatedness and the costs of connecting versus synthesizing

Chihmao Hsieh

Received: 8 April 2010 / Published online: 18 September 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Inventions combine technological features. When features are barely related, burdensomely broad knowledge is required to identify the situations that they share. When features are overly related, burdensomely broad knowledge is required to identify the situations that distinguish them. Thus, according to my first hypothesis, when features are moderately related, the costs of connecting and costs of synthesizing are cumulatively minimized, and the most useful inventions emerge. I also hypothesize that continued experimentation with a specific set of features is likely to lead to the discovery of decreasingly useful inventions; the earlier-identified connections reflect the more common consumer situations. Covering data from all industries, the empirical analysis provides broad support for the first hypothesis. Regressions to test the second hypothesis are inconclusive when examining industry types individually. Yet, this study represents an exploratory investigation, and future research should test refined hypotheses with more sophisticated data, such as that found in literature-based discovery research.

Keywords Connections · Search · Inventions · Patents · Linkage · Relatedness

JEL classification O30 · O32

Introduction

Recent research in research policy and scientometrics has begun investigating the theories by which to predict the value of inventions. Various operationalizations of patent value (Lee 2009; Meyer and Tang 2007) or quality (Cheng et al. 2010; Acosta et al. 2009) notwithstanding, studies have found that past inventors' performance, counts of backward

C. Hsieh (✉)

Amsterdam School of Economics, University of Amsterdam, E2.39, Roetersstraat 11,
1018 WB Amsterdam, The Netherlands
e-mail: c.m.hsieh@uva.nl

citations, and invention size positively correspond to invention value (Bass and Kurgan 2009; Lee 2009; Lee et al. 2007; also see Gay et al. 2005). Such studies help the R&D community to better understand the relationship between technological innovation and their commercial benefits (Lombardo 2008; Thomas 2001), especially insofar that little direct correlation exists between intensity of research activity and the sheer number of patents produced (e.g., Bhattacharya and Meyer 2003).

Yet we know little about any regularities in the manner by which technological features of useful technological inventions are related, such that search for these inventions can be guided. Most notably, Fleming and Sorenson (2001) argue that the inventions of moderate complexity are likely the most useful, striking a balance between ‘fruitful uncertainty and overwhelming complexity’ (Baldwin and Clark 2000: 32). The authors find support for their theory via basic examination of the patent sub-classes associated with a given patent.

While Fleming and Sorenson’s (2001) research begins to guide our understanding of technological usefulness of inventions, a couple difficulties still plague attempts to convert the findings into prescription for search. First, while entrepreneurial or innovation-seeking endeavors often fundamentally depend on search for valuable solutions to valuable problems (Hsieh et al. 2007), Fleming and Sorenson’s (2001) theory doesn’t inform search because it only addresses the ease of combining components from categories with some degree of diversity of other such categories, and in their empirics forgivably the identities of components are eventually washed out. Fleming and Sorenson’s theory simply does not help much in guiding inventors to search for useful *specific* combinations of knowledge. What we lack is some way of accounting for the *relatedness* among knowledge ultimately recombined to create or identify valuable inventions. Second, Fleming and Sorenson’s (2001) empirics do not account for intertemporal changes in the likelihood that classes of components will combine with one another, over time. Changes in the general environment or consumer tastes (e.g., Mackenzie and Wajcman 1985) or advances in science and technology (e.g., Bijker 1987) introduce new sociocultural situations, and the technologies or product features introduced in some industries become better or worse as candidates over time for recombinative endeavors than those in other industries (see Fishman et al. 1993).

I argue that discovery of a valuable invention often requires relating or associating features of a prospective invention to one another via phenomena or principles (Arthur 2007); in other words ‘connections’ exist (see Baron 2006; Baron and Ensley 2006). The more that features can be connected, the higher their relatedness. From this I hypothesize that an inverse U-shaped relationship exists between an invention’s usefulness on one hand and relatedness among its features on the other. I also hypothesize that repeated application or implementation of *specific* sets of features during the evolution of technological advance tends to result in decreasingly useful inventions. The first hypothesis is broadly supported by the empirical results. Indeed, an inverse U-shaped relationship exists between an invention’s usefulness and relatedness among its features, suggesting a tradeoff between minimizing the high costs of connecting unrelated features versus minimizing the high costs of synthesizing highly related ones. Even given these results, however, tests of the second hypothesis are inconclusive.

Below I present the theory, describe the data and methods, and present the results. A discussion describes contributions to the literature as well as limitations of this study, and a conclusion wraps up the paper.

Connections and the measurement of relatedness among an invention's technological features

The design of products or services is often preceded by decision-making regarding the features of the inventions from which products draw (Khilji et al. 2006; Livesay et al. 1989). Here, the term 'feature' is intended to refer to a performance-related aspect (e.g., a visual display's high resolution), a physical part (e.g., a video game system's movement-sensitive controller), or experiential characteristic (e.g., a new age material's softness to touch) of an invention.¹ Choices are made regarding which feature is best. When particular features are considered alongside one another, a series of principles or phenomena—based on means-ends relationships (Shane 2004), shared situations (see Baron 2006), or any combination thereof—may be identified relating or associating them (Arthur 2007). Put another way, a 'connection' has been found (c.f. Baron 2006).

A given set of features may be connected in various ways. A light integrated into an alarm clock helps eliminate clutter in support of bedside nighttime activities. The same light could also be integrated with the alarm clock to shine slowly to a high brightness to subtly awaken its user. In this way a set of features may remain the same (e.g., a light and an alarm clock), even as different connections may be revealed, identified or selected between them (e.g., bedside nighttime activities, or an awakening mechanism).

Relatedness between features is low when few if any connections can be made between them. For example, consider the feature embodied by a fluorescent light, and the feature embodied by a cloth bag. It is not immediately obvious how integrating these two features would be useful. A light is used to illuminate objects or a path in the dark, or to scare away would-be evildoers. Cloth bags are used to carry or cover things. A connection between the two is not immediately obvious.

Yet what at first glance looks like a situation where connections are unavailable between features could just mean that a connection is based on a long series of principles or phenomena. Such 'indirect' connections can be usually reflected by specific demographic market segments, specific occasions, and specific locations. Consider demographic market segments. In contrast to a screen saver program that simply prevents computer screen burn-in, a screen saver program may also relate to a computer screen by providing stimuli that is used to maintain a user's hand-eye coordination or physical well-being, for a specific demographic market segment: senior executives working overtime who must remain sharp to take overseas conference calls. The screen saver and the computer screen itself are thus related in both a direct manner and an indirect one. Alternatively, indirect connections may reflect specific occasions. While the connection between a portable light and a cloth bag is not immediately obvious, one indirect connection can be clearly made in the situation of Halloween, the Western tradition where kids dress up in costumes and walk from door-to-door seeking candy treats. A portable light integrated into the bag's bottom is useful insofar that it helps children to illuminate their path in the dark night to protect themselves not only from obstacles but also from unwelcome strangers leaving one hand free thus not impeding the collection of candy. Finally, specific location may characterize situations indicated by indirect connections. Consider Gamewear, a company that designs and manufactures jewelry combining typical chains, bracelets, and pendants with pieces of sports equipment (Ruth 2006). People like to display sports or team spirit, something that attracts

¹ This definition of 'feature' is my own, elaborating slightly on the definition provided by the American Heritage Dictionary. I use the term 'feature' more generally than 'component,' a term I use that would require some physical part of an invention.

attention and is a ‘conversation starter’ yet does not need to be washed and can easily be put away, which makes it a perfect complement to jewelry. This indirect connection especially applies in the USA and a few other European countries where sports leagues are popular, but would not apply in a country such as Afghanistan, where display of sports team spirit is frowned upon due to religious or political factors (Burns 1996).

Hypotheses

As described above, low relatedness between features is indicated when few connections exist between the features, or in other words when possibly only the most indirect connections are available (i.e., latent). With only the most indirect connections available this suggests that the features are only related in the most specific or specialized situations. Familiarity with these specific and likely rare situations requires either enough luck to ‘be in the right place at the right time,’ or knowledge of various dimensional constraints that delimit the situation. In other words, individuals generally must possess, generate or identify a wide scope of knowledge to valuably *connect* highly unrelated features. Besides requiring broad knowledge, one must also be able to piece together the steps in logic to identify the connection.

If features of an invention are too highly related, then search for useful inventions integrating them is made difficult due to costs of finding the situation that *distinguishes* those features. Those features are more likely to represent similar purposes, uses, or architectures. Particularly useful or valuable integration would then require understanding a wide scope of different theories, phenomena, or social conventions related to increasingly specific situations. The costs of generating an effectively wide scope of knowledge increases as features become overly related. The difficulty shifts from the cost of *connecting* unrelated features to the cost of *synthesizing* overly related ones.

For example, consider an invention that combines today’s rolling shoes—shoes with ‘pop-out’ wheels that allow for rollerskating—with pop-out interlockable planks that can be connected to form a skateboard. In many ways these two are substitutes: they are both used for transportation and exercise in ways that are less jarring than running and more portable than bicycles. To the best of the author’s knowledge, no such invention exists. Why would a skateboarder or rollerblader ever want to pay for a contraption that switches within seconds between the two modes? As illustrated here, it is simply very difficult to synthesize the feature embodied by a pair of rolling shoes, with the feature embodied by a skateboard.

At moderate levels of relatedness among features, valuable connections are more likely discovered. Relatedness between features is not so low that it takes a prohibitively high level of knowledge scope to find situations shared by the features. At the same time, because relatedness is not too high, valuable integration does not require burdensomely costly synthesis. A minimization can be struck between the cost of connecting unrelated features and the cost of synthesizing or integrating overly related ones.

Hypothesis 1 The expected usefulness of inventions is highest when it involves features that tend to be intermediately related to one another.

Past research suggests that the success of a new product or invention is a positive function of the level of the firm’s so-called ‘proximal’ technological experience (Nerkar and Roberts 2004). Experience combining a set of components contributes to a ‘cognitive map’ of the values of solutions to a given problem (Gavetti and Levinthal 2000). Results and lessons learned from experience with specific features can be generalized and

combined with other such lessons to help serve as a map in guiding experimentation with technologically related sets of features (c.f. Fleming and Sorenson 2004; see also Schilling et al. 2003).

However, if we address investment in knowledge (e.g., Dorroh et al. 1994) instead of learning-by-doing (e.g., Adler and Clark 1991), and we measure the effects of repeated experimentation with a *specific* set of features on the usefulness of subsequent inventions using the same sets of features, we are likely to find a more dramatic negative relationship between repeated trials and the usefulness of subsequent inventions. Experience with a set of specific features does little to serve as a map for search over that same set, other than to show which connections have already been tapped. The benefits of generalization described above are no longer relevant. From a more intuitive perspective, the earlier-identified connections among features likely reflect the more commonly occurring situations. Subsequent application or implementation of the particular set of features in future inventions relates to declining usefulness.²

Hypothesis 2 Over time, subsequent application of a particular set of features relates to inventions of declining usefulness.

Empirical considerations

At the heart of both hypotheses is the concept of relatedness, which can be operationalized via any of various candidate methods described in the literature. Some scholars have utilized preset classification schemes to measure relatedness (e.g., Brouthers and Brouthers 2000). While such schemes can appear objective and accurate given the pre-determined codes, the construction of each code's definitional scope and the cumulative set of codes can be highly arbitrary. Also, over time the definition of classes may become distorted if scientific advances render some patent classes or technological areas obsolete, or other definitions simply outdated (e.g., Pavitt 1985 p. 89). Other scholars have developed text-based systems related to "literature-based discovery"³ (seminally, Swanson 1986, 1987) where the typically required input from experts becomes potentially prohibitively costly as analysis extends across industrial or scientific areas. Lastly, still other scholars have investigated the use of citation structure for measuring relatedness across fields, striking a balance between the arbitrariness of classification systems and the costliness of expert-guided text-based systems. Stepping chronologically through citation structure inherently addresses the dynamics of relatedness (i.e., when compared to the matching codes), and also can reasonably approximate a natural path of search.

Of course, the use of citations to measure the origins of knowledge content is often met with skepticism. For example, bibliographic citations (e.g., in academic literature) may pay homage to pioneers, correct the work of others, or criticize that work (Garfield 1962; also see Liu 1993; Bornmann and Daniel 2008). Also, bibliometric citations may reference secondary sources such as literature reviews instead of the seminal paper itself or may leave out citations altogether (MacRoberts and MacRoberts 1989, pp. 343–344). Yet the patent system represents a more rigorous context where citations are made. For a patent to

² For related discussion on efficiencies in experimentation, see Thomke (1998).

³ A subset of literature-based discovery methods includes lexical statistical analysis (Lindsay and Gordon 1999), latent semantic indexing (Gordon and Dumais 1998; Landauer et al. 1998), and association rule mining (Hristovski et al. 2001). For an evaluative review, see Yetisgen-Yildiz and Pratt (2008).

be granted, an innovation must satisfy three requirements: (i) it has to be novel; (ii) non-obvious, in that a skilled practitioner of the technology would not have known how to use it; and (iii) useful. In order to prove novelty, non-obviousness, and usefulness of an invention, the inventor and the patent examiner compare it with prior art through the use of patent citations. Patent references—typically measuring technical knowledge serving as the source of novelty (Sternitzke 2009)—are less likely to be redundant or superfluous than references in journal papers (Collins and Wyatt 1988) due to the controlled nature of the patenting process and its legal consequence (von Wartburg et al. 2005). Additional citations are often added by patent examiners and legal counsel. As Schmoch (1993) explains, because of the novelty requirement the examiner has to look for earlier documents that have the same or almost the same features as the patent application. Only if there are no other relevant documents questioning the novelty of the invention, will the patent application be accepted.

Much past research does agree that patent citations generally reflect features of inventions. As Lanjouw and Schankerman (2001), pp. 133–134 attest: “A patent comprises a set of claims that delineate the boundaries of the property rights provided by the patent. The principal claims define the essential novel features of the invention in their broadest form, and the subordinate claims are more restricted and may describe detailed features of the innovation claimed... Like claims, the citations in the patent document help to define the property rights of the patentee.” Von Wartburg et al. (2005) go onto say that “If two patents are cited, the new invention can be assumed to base equally on both prior patents... The rationale is that the new invention is likely to integrate certain aspects of both former ones, and thus can be regarded as a hybrid development.” On the other hand, these authors also suggest that “the measure of bibliographical coupling is a proxy for the amount of ‘shared-ness’ of technological features among technological variants” (von Wartburg et al. 2005, p. 1599).

Even patent citations can be prohibitively costly from which to draw conclusions. However, compared to European patents, “US patents are more likely to encompass all relevant citations... The US was regarded as a much tougher legal environment. There had to be as much background information as reasonably possible to convince the patent examiner that the prior art was studied closely before filing the application, there were very good distinctions between the claims drawn and the prior art disclosures, and in the event of future litigation there should be good, meaningful distinctions that can be relied upon in a legal battle” (Meyer 2000, p. 108; also see Narin 1994). According to Meyer (2000) p. 106, “US law stipulates that the applicant has to cite any prior art relevant to patentability of the invention known to him or her to the USPTO as long as the application is under examination (‘duty of candor’, USPTO). Non-compliance with this requirement is considered as fraud by the USPTO and can be used as grounds for invalidating the patent.”

Based on the abovementioned considerations, I use data on US patents and both their backwards and forwards citations to test the hypotheses.⁴ Specifically, I construct ‘patent citation networks’ (e.g., Small and Upham 2009; also Milman 1994) to help represent the degree to which features of an invention are related, as measured by the degree to which they have been related in the past. A network of patent citations would reflect connections between features via ‘citation chains’ (see Von Wartburg et al. 2005). Various scholars

⁴ The previous four paragraphs address the reference to a patent’s backward citations in representing the content or derivation of that patent. Of course, one must also consider that not all commercialized or commercializable inventions or innovations are patented, since there exist other appropriation mechanisms—like secrecy or lead time—that industries will differentially use instead (e.g., Levin et al. 1987).

agree. As described by Atallah and Rodriguez (2006) p. 459, “patents can be viewed as elements of a network, with the citations constituting the links between those elements. A patent is linked directly to another patent through a citation, and indirectly through an indirect citation... indirect citations can be of different orders, and hence a patent can be said to be more or less closely related to another patent (and through different channels, i.e., Different citation chains)... a longer chain of citations is indicative of continuity of the impact of an innovation.” von Wartburg et al. 2005 p. 1595 argue that:

...“to map actual developments in a certain technical field and to draw on technological trajectories (Dosi 1982) or avenues (Sahal 1985), citation analysis should rely on everything, bibliographical coupling, co-citations, direct and indirect citations... The technological foundation of citing patents does not only encompass the most recent developments cited directly. It also draws on basic principles provided by earlier patents. Connections to basic patents are revealed by indirect linkages which are captured by citation chains... Given that a patent A cites exclusively patent B which in turn solely cites another patent C, a unique development path can be assumed which stems from C and leads to A.”

Data and methods

I utilize the publicly-available patent database assembled by Hall et al. (2001) that includes information of every invention granted a US patent between 1975 and 1999, a list of all citations that each such invention makes, and a list of all patents that eventually cite it. This database lists each patent’s technological category and subcategories. The usefulness of an invention is measured by future citations, and the network of citations is utilized to measure relatedness. While the database is useful since it accounts for these variables, the empirical analysis requires selecting only the patented inventions where relatedness can be measured among all features.

Dependent variable

Following Fleming and Sorenson (2001) and Narin and Hamilton (1996), I measure the usefulness of inventions via future citation counts over the subsequent 6-year time window (c.f. Maurseth 2005; Wang 2007). Empirical studies have repeatedly shown that future citation counts are indeed related to value (notably see Albert et al. 1991).

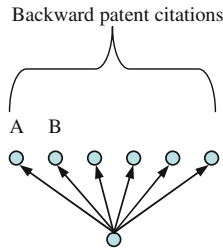
Independent variables

Relatedness

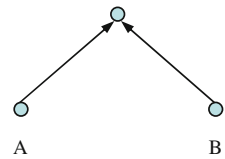
Existing operationalizations of relatedness—e.g., Standard Industrial Classification (SIC) or product codes—are problematic (e.g., Brouthers and Brouthers 2000),⁵ particularly in

⁵ Those methods of determining relatedness that involve measuring patterns of exchange between industries (Burt 1988; Burt and Carlton 1989; Gollop and Monahan 1991; Lemelin 1982), or indices measuring entropy and concentration (i.e., Palepu 1985; Rumelt 1974) are unique to the diversification literature and have little relevance in my context.

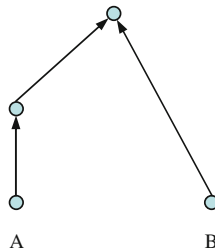
Fig. 1 **a** Shows the relationship between patented inventions and their backward citations. **b–f** Shows the different types of connections between components of an invention. Note that *arrows pointing northward indicate going back in time*. Connection type #2.5 is indicated by either Connection type #2 or Connection type #3



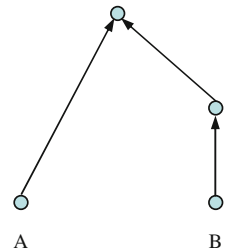
(a) Patented invention



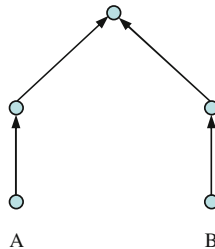
(b) Connection type #1



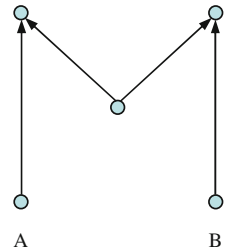
(c) Connection type #2



(d) Connection type #3



(e) Connection type #4



(f) Connection type #5

this setting. Such codes are especially inappropriate in trying to analyze the relatedness of features for which no useful coding scheme can be created. Thus, I measure relatedness among features by examining patent citation histories. One clear advantage to measuring relatedness via patent citation chains instead of via coding schemes is that the actual relatedness of specific features is approximated, instead of relying on coding schemes which only measure the apparent or definitional relatedness at the level of classes of features. Specifically, via a patent citation map (e.g., Huang et al. 2003), I count the number of ways in which two features are connected in the patent citation network. For example two features may be connected in the sense that they both relate to the same backwards citation (Small 1973). As another example they may also be connected as a shared indirect citation; see von Wartburg et al. 2005 pp. 1595–1596. Here the citation

chain is longer than the one in the first example. Five of the different types of citation chains that I examine are shown in Fig. 1.⁶

Utilizing indirect citation chains to help measure relatedness in the empirical analysis imposes additional constraints on the data. Specifically, some inventions are made up of features the relatedness among which cannot be properly measured in the data. Put another way, for some inventions the patent database simply does not go far back enough to catch citation chains that might otherwise have been shown to exist if the citation data had extended farther back in time. Thus, from all 2.14 million inventions granted a US patent between January 1, 1975 and December 31, 1993, I select only those for which all backward citations two generations beforehand were granted after 1975 resulting in a “Patents” dataset corresponding to 18,882 patented inventions. Most patented inventions are based on knowledge fundamentally corresponding to relatively older inventions and of course the requirement of a 6-year window between 1994 through 1999 limits the number of patented inventions that can be examined.⁷

For each of these 18,882 inventions, I take every possible pair of backward citations and treat this as one row. For example if an invention comprises 10 features, it has $10 \times 9/2$ pairs of features and demands 45 rows. Accounting for all 18,882 inventions yields a “Connections” dataset with 318,966 rows.⁸ For each row, I include the frequencies of the different types of citation chains in Fig. 1 that relate the respective pair of features, by running a computerized analysis of the entire 1975–1993 citations list. To calculate indices of relatedness, I merely take the average number of citation chains among all the features of a patent. As an example, consider a hypothetical invention made up of three features A, B, and C. Between features A and B are 3 citation chains of type #1, between B and C lie 5 chains of this type, and between A and C lie 10 such chains. Thus, according to one method of measurement, the relatedness of the invention’s features equates to $(3 + 5 + 10)/3 = 6$. This represents the citation chain type #1 relatedness measure for this invention. I calculate similar statistics for citation chain types #2–#5, and repeat for each of the 18,882 inventions. Finally, because citation chain types #2 and #3 actually represent the same kind of relationship only differing in asymmetry, I create one last measure of relatedness by adding the frequencies of these two citation chain types, and I label this as citation chain type #2.5.^{9,10}

⁶ I explain the derivation of a sixth type later.

⁷ The breakdown of these inventions, by category and subcategory as defined by the patent database, is available from the author upon request.

⁸ Patents citing only one patent are arguably different in nature from those that make two or more citations (c.f. Fleming and Sorenson 2001). During sample selection (described later), patents citing only one patent will be dropped.

⁹ I derived these different types of connections and citation chains. To the best of my knowledge, such a categorization of connections is not extant in the literature.

¹⁰ Linkage bibliographic data (e.g., Brusoni et al. 2005; Callaert et al. 2006; Carpenter and Narin 1983; Meyer 2002; Narin and Noma 1985; Ribeiro et al. 2010; Tijssen et al. 2000)—comprising of patent citations made to the scientific literature and representing the dynamics of the interaction of science and technology (Pavitt 1985; Schmoch 1993, 1997)—are omitted. We are interested here in examining the relatedness of technological features only, not scientific phenomena. And the degree to which scientific literature is cited within patents has been shown to vary by field (e.g., Iversen 2000; Park and Kang 2009; Tamada et al. 2006; Van Looy et al. 2003; Verbeek et al. 2003), which will be controlled in the empirical analysis. We address limitations in the Discussion section.

Measure of prior search along familiar features, number of trials

To estimate the degree that knowledge about a set of features underlying an invention has been accumulated, I determine the number of past technologies involving the exact same set of features, examining the full 1975–1993 portion of the citations database.¹¹

Control variables

Grant date control

This variable helps to control for trends in patenting at the system level (i.e., at the USPTO; see Hall et al. 2001 p. 10). Thus, I add a time variable in terms of number of days after January 1, 1960 that the patent was granted.¹²

Number of components

The number of components (i.e., features) is indicated via backward patent citations. In other words, if features of the invention have already been anticipated by previously existing technologies, any such technologies if patented must be mentioned in the invention's patent application as prior art. Following Schumpeter (1939), all inventions are largely based on recombinations of prior knowledge, and thus backward citations are considered a reasonable measure for a patented invention's features.

Number of classes control

Inventions involving more technological classes are more likely to receive more future citations simply because there are more technological classes that may involve these inventions in the future. This is not unlike how academics who tap into research from various fields are cited by various literatures.

Claims control

Researchers have suggested that the claims made by an invention in a patent application—serving to delineate what is protected by the patent, contingent to patent office approval—signal importance of the invention. Tong and Frame (1994) propose the number of claims as a measure of the 'size' of an innovation, and show that claims-weighted patent counts are more closely related to R&D spending at the national level than simple patent counts.

¹¹ I'd like to thank Jim Hesford for providing the code for this procedure.

¹² Another kind of time-related control variable was considered: the variance in the grant dates of backward citations. According to theory and empirical evidence described by Nerkar (2003), a patented invention would be more valuable if its features derived from a wider range of time, due to the benefits of both temporal exploitation and temporal exploration. I would argue that much of the explanatory power contained in such a variance-oriented variable is already captured in my citation chain frequency statistics: insofar that the rate at which a feature invention is *directly* cited diminishes over time (because its limited *direct* uses are being tapped), there would naturally be fewer citation chains between a feature pair when its underlying features are granted farther apart in time.

Descriptive statistics

Descriptive statistics are shown in Table 1. Most correlations are very low; those that are significant appear to be between variables that derive from one another.

Methods

Because patent citation counts are non-negative, linear regression can yield inefficient, inconsistent, and biased coefficient estimates (Long 1997). Poisson models can be utilized to analyze count data, but they assume that the mean and variance of the observed distribution are equal. Like most count data, the data here exhibit over-dispersion (i.e., the variance exceeds the mean), and negative binomial regressions should be used (see Hausman et al. 1984).

As noted earlier, not all commercialized or commercializable inventions or innovations are patented. Patent statistics underestimate the amount of innovative activity in large firms (Pavitt 1982). Inventors or companies may choose to appropriate value from inventions via secrecy, lead time, learning curve economies, or superior sales and service (Levin et al. 1987). Product inventions and process inventions benefit from these appropriation mechanisms to varying degrees, as do inventions depending on industry type (Levin et al. 1987). Finally, inventors may disclose the minimum necessary depending on the nature of the invention, patent examiners have varying amounts of experience (Cockburn et al. 2003), patent class assignment can be inadequate (as described earlier), and claims can differ in legitimacy depending on industry norms (Merges and Nelson 1990). To begin accounting for this heterogeneity in the propensity to rely on patenting, which can affect conclusions regarding relatedness or usefulness, negative binomial regressions are run separately for each of the six major technological categories as described in the Hall et al. (2001) database: chemical (listed as CAT1), computer and communications (CAT2), drugs and medical CAT3), electrical and electronic (CAT4), mechanical (CAT5), and others (CAT6).

Results

Table 2 shows the results of the negative binomial regressions. Model 1 shows the effects of the control variables. Most notably, the number of distinct technological classes associated with an invention's features is shown to be positively related to future citations. Also, the number of claims an invention makes—often associated to an invention's importance—is positively related to future citations.

The even-numbered subset of Models 2–13 includes the explanatory variables: first- and second-order terms for features, different measures of relatedness, and trials. Hypothesis 1 maintained that usefulness would be highest when relatedness among features is intermediate. In these models where all patented inventions satisfying the data selection constraints are aggregated together from all technological categories, the empirical analysis generally shows that the first-order effect is indeed positive, the second-order effect is negative, and both coefficients are statistically significant. Furthermore, the negative second-order effect does indeed overwhelm the positive first-order effect across the range of relatedness as reported in Table 1. It is not unreasonable to expect that citation chains with the nature of indirectness of type #5 (i.e., the inventions cited by features are linked via subsequent co-citation) should show up as statistically insignificant. Thus, the results provide broad support for H1.

Table 1 Descriptive statistics of variables

Simple correlation coefficients															
Variable	Obs	Mean	Std. Dev.	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Total future citations	18882	4.57	6.44	0	134										
(2) Grant date	18882	35.41	1.80	31.23	37.89	-0.196									
(3) Number of classes	18882	2.00	1.18	1	12	0.071	0.098								
(4) Claims	18882	13.76	11.07	1	197	0.122	0.094	0.081							
(5) Components	18882	5.13	3.55	2	47	0.105	0.100	0.548	0.101						
(6) Components sq	18882	38.92	76.18	4	2209	0.084	0.078	0.472	0.081	0.896					
(7) Connection type 1	18882	0.31	0.68	0	21	-0.006	-0.011	-0.066	-0.019	0.009	0.006				
(8) Connection type 2	18882	0.82	1.95	0	68	0.001	0.004	-0.026	-0.012	0.110	0.081	0.556			
(9) Connection type 3	18882	0.44	1.29	0	67	0.008	-0.002	-0.036	-0.024	0.046	0.033	0.537	0.602		
(10) Connection type 2.5	18882	1.26	2.91	0	105.33	0.004	0.002	-0.033	-0.019	0.094	0.069	0.610	0.936	0.845	
(11) Connection type 4	18882	4.27	13.74	0	407	0.000	-0.005	-0.021	-0.022	0.078	0.059	0.503	0.768	0.658	0.805
(12) Connection type 5	18882	5.04	27.47	0	1854.5	-0.007	0.026	0.004	-0.006	0.055	0.047	0.482	0.496	0.519	0.562
(13) Connection 1 sq	18882	0.55	5.61	0	441	-0.016	0.003	-0.035	-0.019	-0.028	-0.016	0.722	0.305	0.325	0.348
(14) Connection 2 sq	18882	4.46	49.46	0	4624	-0.015	-0.005	-0.019	-0.011	0.019	0.017	0.329	0.740	0.454	0.696
(15) Connection 3 sq	18882	1.86	39.14	0	4489	0.002	-0.006	-0.017	-0.015	-0.010	-0.005	0.177	0.221	0.689	0.453
(16) Connection 2.5 sq	18882	10.06	116.64	0	11095.11	-0.010	-0.004	-0.020	-0.014	0.012	0.011	0.350	0.646	0.657	0.723
(17) Connection 4 sq	18882	207.11	2845.44	0	165649	-0.016	-0.014	-0.016	-0.016	0.010	0.009	0.334	0.539	0.444	0.557
(18) Connection 5 sq	18882	779.74	30456.51	0	3439170	-0.007	0.015	-0.001	0.000	0.003	0.006	0.197	0.206	0.234	0.241
(19) Trials	18882	1.04	0.35	0	13	-0.028	-0.008	-0.050	-0.011	-0.074	-0.079	0.011	-0.010	-0.010	-0.011
(20) Trials sq	18882	1.21	3.10	0	169	-0.011	-0.017	-0.033	-0.002	-0.035	-0.025	-0.003	-0.012	-0.011	-0.013

Table 1 continued

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
(12) Connection type 5	0.505								
(13) Connection 1 sq	0.335	0.360							
(14) Connection 2 sq	0.615	0.411	0.269						
(15) Connection 3 sq	0.290	0.260	0.148	0.313					
(16) Connection 2.5 sq	0.602	0.467	0.303	0.910	0.641				
(17) Connection 4 sq	0.827	0.407	0.298	0.699	0.295	0.690			
(18) Connection 5 sq	0.222	0.789	0.209	0.301	0.194	0.368	0.291		
(19) Trials	-0.009	-0.008	0.003	0.002	-0.002	-0.001	0.003	-0.002	
(20) Trials sq	-0.009	-0.007	0.000	-0.001	-0.002	-0.002	-0.001	-0.001	0.901

Each observation refers to an invention, and thus connection-related variables here relate to averages over all pairs of an invention's components. "sq" indicates the square of the term

Table 2 Negative binomial regressions of future citations

Controls	Connection type #1 $X = 1$			Connection type #2 $X = 2$			Connection type #3 $X = 3$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Grant date	-0.194*** (0.0052)	-0.200*** (0.0052)	-0.200*** (0.0052)	-0.201*** (0.0052)	-0.200*** (0.0052)	-0.201*** (0.0052)	-0.200*** (0.0052)	-0.200*** (0.0052)	-0.200*** (0.0052)
Number of classes	0.0822*** (0.0068)	0.0142 (0.0083)	0.0151 (0.0083)	0.0155 (0.0084)	0.0162 (0.0084)	0.0149 (0.0083)	0.0157 (0.0083)	0.0157 (0.0083)	0.0157 (0.0083)
Claims	0.0166*** (0.0008)	0.0159*** (0.0008)	0.0159*** (0.0008)	0.0159*** (0.0008)	0.0159*** (0.0008)	0.0160*** (0.0008)	0.0160*** (0.0008)	0.0160*** (0.0008)	0.0160*** (0.0008)
Components	0.0665*** (0.0054)	0.0679*** (0.0054)	0.0679*** (0.0054)	0.0650*** (0.0055)	0.0664*** (0.0055)	0.0664*** (0.0054)	0.0679*** (0.0054)	0.0679*** (0.0054)	0.0679*** (0.0054)
Components sq	-0.00135*** (0.0002)	-0.00135*** (0.0002)	-0.00148*** (0.0002)	-0.00132*** (0.0002)	-0.00145*** (0.0002)	-0.00135*** (0.0002)	-0.00148*** (0.0002)	-0.00148*** (0.0002)	-0.00148*** (0.0002)
Connection type X	0.0281 (0.0191)	0.0319 (0.0192)	0.0319 (0.0192)	0.0283*** (0.0082)	0.0287*** (0.0082)	0.0195* (0.0092)	0.0200* (0.0092)	0.0200* (0.0092)	0.0200* (0.0092)
Connection X sq	-0.00650* (0.0027)	-0.00679* (0.0027)	-0.00679* (0.0027)	-0.00220*** (0.0005)	-0.00220*** (0.0005)	-0.000238 (0.0002)	-0.000246 (0.0002)	-0.000246 (0.0002)	-0.000246 (0.0002)
Trials	-0.117* ** (0.0219)	-0.117* ** (0.0219)	-0.117* ** (0.0219)	-0.115*** (0.0219)	-0.115*** (0.0219)	-0.116*** (0.0219)	-0.116*** (0.0219)	-0.116*** (0.0219)	-0.116*** (0.0219)
Trials sq	0.0295*** (0.0067)	0.0295*** (0.0067)	0.0295*** (0.0067)	0.0295*** (0.0067)	0.0295*** (0.0067)	0.0295*** (0.0067)	0.0295*** (0.0067)	0.0295*** (0.0067)	0.0295*** (0.0067)
Constant	7.932*** (0.1820)	8.113*** (0.1840)	8.314*** (0.1880)	8.129*** (0.1830)	8.329*** (0.1880)	8.122*** (0.1830)	8.323*** (0.1880)	8.323*** (0.1880)	8.323*** (0.1880)
Lnalpha constant	0.0821*** (0.0131)	0.0638*** (0.0132)	0.0624*** (0.0132)	0.0628*** (0.0132)	0.0613*** (0.0132)	0.0641*** (0.0132)	0.0626*** (0.0132)	0.0626*** (0.0132)	0.0626*** (0.0132)
N	18,882	18,882	18,882	18,882	18,882	18,882	18,882	18,882	18,882
Log likelihood	-48462.553	-48327.647	-48316.101	-48318.381	-48307.065	-48328.356	-48316.915	-48316.915	-48316.915

Table 2 continued

Controls Model	Connection type #2.5 $X = 2.5$			Connection type #4 $X = 4$			Connection type #5 $X = 5$		
	(8)	(9)	(10)	(11)	(12)	(13)	(12)	(13)	(13)
Grant date	-0.201*** (0.0052)	-0.200*** (0.0052)	-0.201*** (0.0052)	-0.201*** (0.0052)	-0.200*** (0.0052)	-0.200*** (0.0052)	-0.200*** (0.0052)	-0.200*** (0.0052)	-0.200*** (0.0052)
Number of classes	0.0157 (0.0084)	0.0165* (0.0084)	0.0166* (0.0083)	0.0174* (0.0083)	0.0133 (0.0083)	0.0133 (0.0083)	0.0133 (0.0083)	0.0133 (0.0083)	0.0133 (0.0083)
Claims	0.0160*** (0.0008)	0.0159*** (0.0008)	0.0160*** (0.0008)	0.0160*** (0.0008)	0.0160*** (0.0008)	0.0160*** (0.0008)	0.0160*** (0.0008)	0.0160*** (0.0008)	0.0160*** (0.0008)
Components	0.0653*** (0.0054)	0.0667*** (0.0055)	0.0647*** (0.0054)	0.0662*** (0.0054)	0.0671*** (0.0054)	0.0671*** (0.0054)	0.0671*** (0.0054)	0.0671*** (0.0054)	0.0671*** (0.0054)
Components sq	-0.00133*** (0.0002)	-0.00145*** (0.0002)	-0.00132*** (0.0002)	-0.00145*** (0.0002)	-0.00136*** (0.0002)	-0.00136*** (0.0002)	-0.00136*** (0.0002)	-0.00136*** (0.0002)	-0.00136*** (0.0002)
Connection type X	0.0130** (0.0047)	0.0133** (0.0047)	0.00595*** (0.0012)	0.00591*** (0.0012)	0.000192 (0.0006)	0.000192 (0.0006)	0.000192 (0.0006)	0.000192 (0.0006)	0.000192 (0.0006)
Connection X sq	-0.000430** (0.0001)	-0.000435*** (0.0001)	0.0000411*** (0.0000)	0.0000409*** (0.0000)	-6.55E-07 (0.0000)	-6.55E-07 (0.0000)	-6.55E-07 (0.0000)	-6.55E-07 (0.0000)	-6.55E-07 (0.0000)
Trials	-0.116*** (0.0219)	-0.367*** (0.0572)	-0.115*** (0.0219)	-0.363*** (0.0572)	-0.116*** (0.0219)	-0.116*** (0.0219)	-0.116*** (0.0219)	-0.116*** (0.0219)	-0.116*** (0.0219)
Trials sq		0.0294*** (0.0067)		0.0290*** (0.0067)		0.0290*** (0.0067)		0.0290*** (0.0067)	
Constant	8.129*** (0.1830)	8.330*** (0.1880)	8.142*** (0.1830)	8.340*** (0.1880)	8.122*** (0.1840)	8.122*** (0.1840)	8.122*** (0.1840)	8.122*** (0.1840)	8.122*** (0.1840)
Lnalpha co ns tant	0.0636*** (0.0132)	0.0622*** (0.0132)	0.0619*** (0.0132)	0.0605*** (0.0132)	0.0642*** (0.0132)	0.0642*** (0.0132)	0.0642*** (0.0132)	0.0642*** (0.0132)	0.0642*** (0.0132)
N	18,882	18,882	18,882	18,882	18,882	18,882	18,882	18,882	18,882
Log likelihood	-48325.2	-48313.703	-48311.712	-48300.536	-48330.045	-48330.045	-48330.045	-48330.045	-48330.045

Note that the type of connection used to measure relatedness and the squared term is indicated by column. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Hypothesis 2 maintained that future attempts to combine a previously combined set of specific features ultimately lead to inventions of lower usefulness. The even-numbered models of Table 2 show a negative relationship between the number of past trials with a set of features, and the usefulness of an invention that uses such a set. In this preliminary analysis, the most useful configurations for a given set of features appear typically found first.

I include a second-order term for number of trials. As the odd-numbered subset of Models 3–13 shows, there appears to be a nonlinear relationship between number of past trials and usefulness. At first, number of trials is negatively related to usefulness. As experience with a set of features accumulates, the negative relationship begins to disappear.

Since it is well-accepted that the propensity to patent inventions differs across industries (Cohen et al. 2000; Levin et al. 1987), additional negative binomial regressions are run but this time with category fixed effects. Table 3 supports the notion that some industries are more likely to patent than others.¹³ Specifically, the computer and electronics industries apparently have a greater tendency to patent than the chemical and drug industries. Whether these coefficients indicate differences in rates of innovation or differences in rates of patenting vis-a-vis other appropriation mechanisms is unknown from this data alone (c.f. Pavitt 1982), and would have to be combined with other variables indicating innovation-based inputs or innovation output measures. More importantly, even after controlling for industry via these fixed effects, Hypothesis 1 is still supported.

Table 4 shows negative binomial regressions for each technological category. Hypothesis 1 linking an invention's usefulness and the relatedness among its features is supported fully for three of the five definitive technological categories (computer-, electronics-, and mechanically-oriented industries). However, the relationship between number of prior trials and invention usefulness found in Table 2 is generally not evidenced by these industry-specific regressions. Thus, when industry-specific regressions are run, Hypothesis 2 is unsupported.

Discussion

The paper is the first to my knowledge to explicitly examine the relationship between an invention's usefulness and the socioculturally oriented relatedness of its features. Generally speaking, a statistically significant inverse U-shaped relationship is found between an invention's usefulness and the relatedness among its features, evident from models that measure relatedness by citation chain types #2, #2.5, and #4. Connections between features too direct do not seem related to the usefulness of a focal invention (i.e., citation chain type #1), and the same appears to hold for connections based on linking inventive features after they have been discovered (i.e., citation chain type #5).

The data used to test the hypotheses has a couple limitations. First, the data likely does not account for all of any given invention's features; non-patent citations are not covered. While the focus of this paper is indeed on technological features and development (e.g., Verbeek et al. 2003) and not scientific phenomena underlying patented inventions, the structure underlying the connectedness among the patent citations (i.e., technological features) can help to indicate the scientific phenomena at hand (c.f. Faucompré et al. 1997; Lo 2010; Narin and Noma 1985). To re-iterate, citations made to patents have also been

¹³ Tables 3 and 4 only measure relatedness based on citation chain type #4. There are no substantive differences in statistical analysis when re-calculating for citation chain types #2 or the more important #2.5.

Table 3 Negative binomial regression of future citations, with category fixed effects

Model	1	2	3	4
Grant date	-0.223*** (0.005)	-0.223*** (0.005)	-0.226*** (0.005)	-0.226*** (0.005)
Number of classes	0.0239** (0.008)	0.0241** (0.008)	0.0163* (0.008)	0.0164* (0.008)
Claims	0.0152*** (0.001)	0.0152*** (0.001)	0.0146*** (0.001)	0.0146*** (0.001)
Components	0.0445*** (0.005)	0.0452*** (0.005)	0.0385*** (0.005)	0.0388*** (0.005)
Components sq	-0.000819*** (0.000)	-0.000876*** (0.000)	-0.000669** (0.000)	-0.000691** (0.000)
Relatedness	0.206*** (0.034)	0.207*** (0.034)	0.150*** (0.034)	0.150*** (0.034)
Relatedness sq	-0.0389*** (0.007)	-0.0390*** (0.007)	-0.0306*** (0.007)	-0.0307*** (0.007)
Trials	-0.0414 (0.023)	-0.144* (0.056)	-0.0241 (0.024)	-0.0643 (0.056)
Trials sq		0.0126 (0.006)		0.00495 (0.006)
CAT1	-0.204*** (0.038)	-0.202*** (0.038)		
CAT2	0.679*** (0.030)	0.679*** (0.030)		
CAT3	-0.250*** (0.038)	-0.245*** (0.038)		
CAT4	0.341*** (0.031)	0.341*** (0.031)		
CAT6	-0.0463 (0.053)	-0.0442 (0.053)		
Subcategory fixed effects			Not statistically significant	
Constant	8.593*** (0.177)	8.673*** (0.182)	9.686*** (1.023)	9.718*** (1.024)
Lnalpha constant	-0.0568*** (0.014)	-0.0570*** (0.014)	-0.0976*** (0.014)	-0.0976*** (0.014)
N	18,882	18,882	18,882	18,882
Log likelihood	-47467.26	-47465.21	-47174.86	-47174.54

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Table 4 Negative binomial regression of future citations, per technological category

Model	Chemical CAT = 1			Comp and Comm CAT = 2			Drugs and Medical CAT = 3		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Grant date	-0.222*** (0.0169)	-0.222*** (0.0169)	-0.220*** (0.0086)	-0.221*** (0.0086)	-0.283*** (0.0187)	-0.283*** (0.0187)			
Number of classes	0.0786** (0.0280)	0.0769** (0.0280)	0.0223 (0.0116)	0.0224 (0.0116)	0.0526 (0.0383)	0.0524 (0.0383)			
Claims	0.0216*** (0.0030)	0.0217*** (0.0030)	0.0174*** (0.0012)	0.0174*** (0.0012)	0.0138*** (0.0025)	0.0137*** (0.0025)			
Components	0.0434* (0.0172)	0.0519** (0.0175)	0.0404*** (0.0081)	0.0438*** (0.0090)	0.0749* (0.0325)	0.0747* (0.0325)			
Components sq	-0.00104 (0.0006)	-0.00161* (0.0006)	-0.000636* (0.0003)	-0.000851* (0.0004)	-0.00322 (0.0020)	-0.00321 (0.0020)			
Relatedness	-0.00086 (0.0024)	-0.000923 (0.0024)	0.00860*** (0.0022)	0.00857*** (0.0022)	-0.00508 (0.0058)	-0.00511 (0.0058)			
Relatedness sq	-0.00000689 (0.0000)	-0.00000662 (0.0000)	-0.0000376* (0.0000)	-0.0000374* (0.0000)	0.00000781 (0.0000)	0.00000791 (0.0000)			
Trials	-0.418*** (0.1100)	-1.372** (0.4920)	0.00693 (0.0984)	-0.411 (0.4720)	0.0168 (0.0417)	-0.0182 (0.1150)			
Trials sq		0.260* (0.1310)		0.134 (0.1490)		0.00438 (0.0136)			
Constant	8.540*** (0.5900)	9.243*** (0.6890)	9.065*** (0.3170)	9.343*** (0.4420)	10.28*** (0.6610)	10.31*** (0.6710)			
Lnalpha constant	0.225*** (0.0474)	0.223*** (0.0474)	-0.0638** (0.0203)	-0.0639** (0.0203)	0.381*** (0.0475)	0.381*** (0.0475)			
N	1,965	1,965	7,018	7,018	2,033	2,033			
Log likelihood	-4131.7333	-4129.767	-19811.652	-19811.231	-4019.2834	-4019.2309			
Model	Electrical and Electronic CAT = 4			Mechanical CAT = 5			Others CAT = 6		
	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Grant date	-0.222*** (0.0086)	-0.222*** (0.0086)	-0.193*** (0.0129)	-0.194*** (0.0129)	-0.208*** (0.0269)	-0.209*** (0.0269)			
Number of classes	-0.0227 (0.0162)	-0.0224 (0.0162)	0.0277 (0.0267)	0.0274 (0.0267)	0.130* (0.0529)	0.130* (0.0529)			
Claims	0.0131*** (0.0014)	0.0131*** (0.0014)	0.00727*** (0.0021)	0.00726*** (0.0021)	0.0144*** (0.0046)	0.0144*** (0.0046)			
Components	0.0665*** (0.0112)	0.0664*** (0.0112)	0.0343 (0.0187)	0.0213 (0.0207)	0.048 (0.0395)	0.0479 (0.0395)			
Components sq	-0.00170** (0.0006)	-0.00170** (0.0006)	-0.000882 (0.0010)	0.0000494 (0.0012)	-0.00252 (0.0021)	-0.00251 (0.0021)			
Relatedness	0.0106*** (0.0027)	0.0106*** (0.0027)	0.0177*** (0.0035)	0.0178*** (0.0035)	0.0231* (0.0116)	0.0232* (0.0116)			
Relatedness sq	-0.0000537** (0.0000)	-0.0000539** (0.0000)	-0.0000955** (0.0000)	-0.0000965** (0.0000)	-0.000196 (0.0001)	-0.000197 (0.0001)			
Trials	-0.0416 (0.0352)	-0.107 (0.1080)	-0.287 (0.1840)	1.345 (1.1400)	-0.264 (0.1690)	0.0907 (0.8470)			

Table 4 continued

Model	Electrical and Electronic CAT = 4		Mechanical CAT = 5		Others CAT = 6	
	(7)	(8)	(9)	(10)	(11)	(12)
Trials sq		0.00612 (0.0096)		−0.534 (0.3700)		−0.0986 (0.2300)
Constant	8.864*** (0.3060)	8.925*** (0.3200)	7.861*** (0.4820)	6.814*** (0.8630)	8.021*** (0.9690)	7.770*** (1.1320)
Lnalpha constant	−0.167*** (0.0255)	−0.167*** (0.0255)	−0.423*** (0.0521)	−0.424*** (0.0522)	−0.183* (0.0855)	−0.184* (0.0855)
N	5,441	5,441	1,790	1,790	635	635
Log likelihood	−13857.401	−13857.198	−4081.8651	−4080.7912	−1411.395	−1411.303

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

shown to indicate novelty, in a qualitatively different fashion compared to citations made to publications (also see Meyer 2000). Second, the data may account for features that have little to do with the invention itself. Specifically, some backwards citations that become listed as features may be cited extraneously as substitutes for other features, for the sole purpose of documentation.

This paper helps respond to recent work lamenting that the effects of relatedness are not well-understood. As described by D'aveni et al. (2004) pp. 365–366, "...the empirical search for synergistic effects (from resource-sharing among related businesses) on corporate-level performance has produced mixed and inconsistent results... Mixed results suggest that scholars need to understand the impact of diversification at a finer-grained level of detail (Lubatkin et al. 2001)." And while prior work on resource-based synergies has focused on equating synergy to mere relatedness and economies of scale or scope (e.g., Gary 2005; Schilling et al. 2003; St. John and Harrison 1999; for earlier work, see Amit and Livnat 1988; Barney 1988; Davis and Thomas 1993; and their references), the current paper specifically examines relatedness at a finer-grained dynamic level. Instead of examining relatedness according to arbitrarily assigned coding schemes, this paper examines relatedness according to a patent system that requires records of association between features regardless of coding schemes (in other words, patent examiners are generally understood to be familiar enough with the prior art that they will include all relevant patent citations as prior art). Of course a measure of relatedness based on informetric or citation-based analysis may still suffer from the arbitrariness of coding schemes, insofar that pieces of knowledge (i.e., patents) are categorized by codes, and inventors are expected to search and cite knowledge from some fields more thoroughly than others (i.e., by patent examiners). However, insofar that inventors are required to (or require themselves to) search for all relevant prior art regardless of those codes, then operationalizations of relatedness based on citation-based analysis may be particularly appropriate.

Finally, the findings also reveal a handful of relatively novel future research questions or directions, at the more fine-grained level. First, research can be done to refine understanding of what else moderates the effects of prior experience on the search for useful inventions. For example one might expect that repeated trials and innovations exploring familiar features should lead to greater marginal improvements in usefulness when inventors update their own personal cognitive representations, versus adjusting cognitive representations due to the prior efforts of others. More nuanced data for testing would also be preferred. Second, future research could examine more of the types of connections created between features after those features have been discovered but before the invention has been granted patent rights (e.g., citation chain type #5). Third, successfully relating patent citation networks to the usefulness of invention (underlying products or opportunities) may clue researchers into the nature of search processes. While Fleming and Sorenson's (2001) analysis leaves some room for interpreting how search specifically might take place, this study posits that individuals may systematically search across knowledge spaces by examining specific features or other inventions that have used those features, much like the process that academics use upon analyzing bibliographies, the 'Web of Science,' or 'Google Scholar.'

Conclusion

This study is one of the first attempting to identify links between an invention's usefulness and two variables: the degree of relatedness among its features, and the number of times

the invention's specific set of features has been used for prior inventions. We simply know little about any regularities in the manner by which technological features of useful technological inventions are related, such that search for these inventions can be guided. As argued, when features are barely related, burdensomely broad knowledge is required to identify the situations that they share since the features are only related in relatively specific or specialized situations. When features are overly related, burdensomely broad knowledge is required to identify the specific situations that distinguish them. When features are moderately related, the costs of connecting and costs of synthesizing are cumulatively minimized, and the most useful inventions emerge. I also hypothesize that continued experimentation with a specific set of features is likely to lead to the discovery of decreasingly useful inventions; the earlier-identified connections reflect the more common consumer situations. Covering data from all industries, the empirical analysis provides broad support for only the first hypothesis. Regressions to test the second hypothesis are less conclusive, however, when examining industry types individually.

Besides using patent citation data to investigate the determinants of invention usefulness, this paper also preliminarily explores an operationalization of relatedness at a fine-grained dynamic level. While patent data is nuanced enough to reflect technological connections or linkages, as suggested by previous scholars (e.g., von Wartburg et al. 2005), raw patent citation data does not appear to provide clues to the cross-time changes in usefulness of inventions when they make the exact same set of prior patent citations. As information processing technologies advance, future research may be able to test refined hypotheses across industries with sophisticated data analysis related to literature-based discovery.

Acknowledgments This paper derives from my dissertation. I wish to thank Sergio Lazzarini, and also my committee at Washington University's Olin School of Business (especially Todd Zenger and Jackson Nickerson) for suggestions and comments on prior versions. All errors remain mine.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Acosta, M., Coronado, D., & Fernández, A. (2009). Exploring the quality of environmental technology in Europe: Evidence from patent citations. *Scientometrics*, 80(1), 131–152.
- Adler, P., & Clark, K. (1991). Behind the learning curve: A sketch of the learning process. *Management Science*, 37(3), 267–281.
- Albert, M., Avery, D., Narin, F., & McAllister, P. (1991). Direct validation of citation counts as indicators of industrially important patents. *Research Policy*, 20(3), 251–259.
- Amit, R., & Livnat, J. (1988). Diversification strategies, business cycles, and economic performance. *Strategic Management Journal*, 9, 99–110.
- Arthur, W. B. (2007). The structure of invention. *Research Policy*, 36(2), 274–287.
- Atallah, G., & Rodriguez, G. (2006). Indirect patent citations. *Scientometrics*, 67(3), 437–465.
- Baldwin, C., & Clark, K. (2000). *Design rules: The power of modularity*. Cambridge, MA: MIT Press.
- Barney, J. (1988). Returns to bidding firms in mergers and acquisitions: Reconsidering the relatedness hypothesis. *Strategic Management Journal*, 9, 71–78.
- Baron, R. (2006). Opportunity recognition as pattern recognition: How entrepreneurs 'connect the dots' to identify new business opportunities. *Academy of Management Perspectives*, 20(1), 104–119.
- Baron, R., & Ensley, M. (2006). Opportunity recognition as the detection of meaningful patterns: Evidence from comparisons of novice and experienced entrepreneurs. *Management Science*, 52(9), 1331–1344.

- Bass, S. D., & Kurgan, L. A. (2009). Discovery of factors influencing patent value based on machine learning in patents in the field of nanotechnology. *Scientometrics*, 82(2), 217–241.
- Bhattacharya, S., & Meyer, M. (2003). Large firms and the science-technology interface: Patents, patent citations, and scientific output of multinational corporations in thin films. *Scientometrics*, 58(2), 265–279.
- Bijker, W. (1987). The social construction of Bakelite: Toward a theory of invention. In W. Bijker, T. Hughes, & T. Pinch (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology* (pp. 159–187). Cambridge, MA: MIT Press.
- Bornmann, L., & Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45–80.
- Brouthers, K., & Brouthers, L. (2000). Acquisition or greenfield start-up? Institutional, cultural and transaction cost influences. *Strategic Management Journal*, 21(1), 89–97.
- Brusoni, S., Criscuolo, P., & Geuna, A. (2005). The knowledge bases of the world's largest pharmaceutical groups: What do patent citations to non-patent literature reveal? *Economics of Innovation and New Technology*, 14(5), 395–415.
- Burns, J. (1996, December 19). At an Afghan execution: It's swift and personal. *The New York Times*, electronic edition.
- Burt, R. (1988). The stability of American markets. *American Journal of Sociology*, 94, 356–395.
- Burt, R., & Carlton, D. (1989). Another look at the network boundaries of American markets. *American Journal of Sociology*, 75, 723–753.
- Callaert, J., Van Looy, B., Verbeek, A., Debackere, K., & Thijs, B. (2006). Traces of prior art: An analysis of non-patent references found in patent documents. *Scientometrics*, 69(1), 3–20.
- Carpenter, M. P., & Narin, F. (1983). Validation study: Patent citations as indicators of science and foreign dependence. *World Patent Information*, 5, 180–185.
- Cheng, Y.-H., Kuan, F.-Y., Chuang, S.-C., & Ken, Y. (2010). Profitability decided by patent quality? An empirical study of the U.S. semiconductor industry. *Scientometrics*, 82(1), 175–183.
- Cockburn, I., Kortum, S., & Stern, S. (2003). Are all patent examiners equal? Examiners, patent characteristics, and litigation outcomes. In W. Cohen & S. Merrill (Eds.), *Patents in the knowledge-based economy* (pp. 19–53). Washington, DC: National Academies Press.
- Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2000). Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not). NBER Working Paper #7552.
- Collins, P., & Wyatt, S. (1988). Citations in patents to the basic research literature. *Research Policy*, 17, 65–74.
- D'Aveni, R. A., Ravenscraft, D. J., & Anderson, P. (2004). From corporate strategy to business-level advantage: Relatedness as resource congruence. *Managerial and Decision Economics*, 25(6–7), 365–381.
- Davis, R., & Thomas, L. (1993). Direct estimation of synergy: A new approach to the diversity-performance debate. *Management Science*, 39(11), 1334–1346.
- Dorroh, J., Gulledge, T., & Womer, N. (1994). Investment in knowledge: A generalization of learning by experience. *Management Science*, 40(8), 947–958.
- Dosi, G. (1982). Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. *Research Policy*, 22(2), 102–103.
- Faucompré, P., Quoniam, L., & Dou, H. (1997). An effective link between science and technology. *Scientometrics*, 40(3), 465–480.
- Fishman, A., Gandal, N., & Shy, O. (1993). Planned obsolescence as an engine of technological progress. *Journal of Industrial Economics*, 41(4), 361–370.
- Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: Evidence from patent data. *Research Policy*, 30(7), 1019–1039.
- Fleming, L., & Sorenson, O. (2004). Science as a map in technological search. *Strategic Management Journal*, 25, 909–928.
- Garfield, E. (1962). Can citation indexing be automated? *Essays of an Information Scientist*, 1, 84–90.
- Gary, M. (2005). Implementation strategy and performance outcomes in related diversification. *Strategic Management Journal*, 26, 643–664.
- Gavetti, G., & Levinthal, D. (2000). Looking forward and looking backward: Cognitive and experiential search. *Administrative Science Quarterly*, 45(1), 113–139.
- Gay, C., LeBas, C., Patel, P., & Touach, K. (2005). The determinants of patent citations: An empirical analysis of French and British patents in the US. *Economics of Innovation and New Technology*, 14(5), 339–350.
- Gollop, F. M., & Monahan, J. L. (1991). A generalized index of diversification: trends in US manufacturing. *Review of Economics and Statistics*, 73, 318–330.
- Gordon, M. D., & Dumais, S. (1998). Using latent semantic indexing for literature based discovery. *Journal of the American Society for Information Science*, 49(8), 674–685.

- Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). The NBER patent citation data file: Lessons, insights and methodological tools. NBER Working Paper #8498.
- Hausman, J., Hall, B. H., & Griliches, Z. (1984). Econometric models for count data with an application to the patents-R&D relationship. *Econometrica*, 52(4), 909–938.
- Hristovski, D., Stare, J., Peterlin, B., & Dzeroski, S. (2001). Supporting discovery in medicine by association rule mining in Medline and UMLS. *Studies in Health Technology and Informatics*, 84(pt 2), 1344–1348.
- Hsieh, C. M., Nickerson, J. A., & Zenger, T. R. (2007). Opportunity discovery, problem solving, and the entrepreneurial theory of the firm. *Journal of Management Studies*, 44(7), 1255–1277.
- Huang, M.-H., Chiang, L.-Y., & Chen, D.-Z. (2003). Constructing a patent citation map using bibliographic coupling: A study of Taiwan's high-tech companies. *Scientometrics*, 58(3), 489–506.
- Iversen, E. J. (2000). An excursion into the patent-bibliometrics of Norwegian patenting. *Scientometrics*, 49(1), 63–80.
- Khilji, S. E., Mroczkowski, T., & Bernstein, B. (2006). From invention to innovation: Toward developing an integrated innovation model for biotech firms. *Journal of Product Innovation Management*, 23(6), 528–540.
- Landauer, T., Foltz, P., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, 25, 259–284.
- Lanjouw, J., & Schankerman, M. (2001). Characteristics of patent litigation: A window on competition. *RAND Journal of Economics*, 32(1), 129–151.
- Lee, Y.-G. (2009). What affects a patent's value? An analysis of variables that affect technological, direct economic, and indirect economic value: An exploratory conceptual approach. *Scientometrics*, 79(3), 623–633.
- Lee, Y.-G., Lee, J.-D., Song, Y.-I., & Lee, S.-J. (2007). An in-depth empirical analysis of patent citation counts using zero-inflated count data model: The case of KIST. *Scientometrics*, 70(1), 27–39.
- Lemelin, A. (1982). Relatedness in the patterns of inter-industry diversification. *Review of Economics and Statistics*, 64, 646–657.
- Levin, R., Klevorick, A., Nelson, R., & Winter, S. (1987). Appropriating the returns from industrial research and development: Comments and discussion. *Brookings Papers on Economic Activity*, 3, 783–831.
- Lindsay, R. K., & Gordon, M. D. (1999). Literature based discovery by lexical statistics. *Journal of the American Society for Information Science*, 50(7), 574–587.
- Liu, M. (1993). Progress in documentation—the complexities of citation practice: A review of citation studies. *Journal of Documentation*, 49, 370–408.
- Livesay, H. C., Rorke, M. L., & Lux, D. S. (1989). Technical development and the innovation process. *Journal of Product Innovation Management*, 6(4), 268–281.
- Lo, S.-c. S. (2010). Scientific linkage of science research and technology development: A case of genetic engineering research. *Scientometrics*, 82, 109–120.
- Lombardo, L. (2008). New indicators linking patenting and business R&D expenditure. *Scientometrics*, 76(2), 201–224.
- Long, J. (1997). *Modeling frequency and count data*. Oxford: Oxford University Press.
- Lubatkin, M., Schulze, W., Mainkar, A., & Cotterill, R. (2001). Ecological investigation of firm effects in horizontal mergers. *Strategic Management Journal*, 22, 335–357.
- MacKenzie, D., & Wajcman, J. (Eds.). (1985). *The social shaping of technology*. Philadelphia, PA: Open University Press.
- MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science*, 40, 342–349.
- Maurseth, P. B. (2005). Lovely but dangerous: The impact of patent citations on patent renewal. *Economics of Innovation and New Technology*, 14(5), 351–374.
- Merges, R. P., & Nelson, R. R. (1990). On the complex economies of patent scope. *Columbia Law Review*, 90, 839–916.
- Meyer, M. (2000). What is special about patent citations? Differences between scientific and patent citations. *Scientometrics*, 49(1), 93–123.
- Meyer, M. (2002). Tracing knowledge flows in innovation systems. *Scientometrics*, 54(2), 193–212.
- Meyer, M. S., & Tang, P. (2007). Exploring the “value” of academic patents: IP management practices in UK universities and their implications for Third-Stream indicators. *Scientometrics*, 70(2), 415–440.
- Milman, B. L. (1994). Individual co-citation clusters as nuclei of complete and dynamic informetric models of scientific and technological areas. *Scientometrics*, 31(1), 45–57.
- Narin, F. (1994). Patent bibliometrics. *Scientometrics*, 30(1), 147–155.
- Narin, F., & Hamilton, K. S. (1996). Bibliometric performance measures. *Scientometrics*, 36(3), 293–310.
- Narin, F., & Noma, E. (1985). Is technology becoming science? *Scientometrics*, 7, 369–381.

- Nerkar, A. (2003). Old is gold? The value of temporal exploration in the creation of new knowledge. *Management Science*, 49(2), 211–229.
- Nerkar, A., & Roberts, P. W. (2004). Technological and product-market experience and the success of new product introductions in the pharmaceutical industry. *Strategic Management Journal*, 25, 779–799.
- Palepu, K. (1985). Diversification strategy, profit performance and the entropy measure. *Strategic Management Journal*, 6, 239–255.
- Park, H. W., & Kang, J. (2009). Patterns of scientific and technological knowledge flows based on scientific papers and patents. *Scientometrics*, 81(3), 811–820.
- Pavitt, K. (1982). R&D, patenting and innovative activities. *Research Policy*, 11, 33–51.
- Pavitt, K. (1985). Patent statistics as indicators of innovative activities: Possibilities and problems. *Scientometrics*, 7(1–2), 77–99.
- Ribeiro, L. C., Ruiz, R. M., Bernardes, A. T., & Albuquerque, E. M. (2010). Matrices of science and technology interactions and patterns of structured growth: Implications for development. *Scientometrics*, 83(1), 55–75.
- Rumelt, R. P. (1974). *Strategy, structure, and economic performance*. Cambridge, MA: Harvard University Press.
- Ruth, J.-P. S. (2006, April 24). GameWear lets fans wear their heroes home. *NJBIZ*, 19, 3–4.
- Sahal, D. (1985). Technological guideposts and innovation avenues. *Research Policy*, 14, 61–82.
- Schilling, M., Vidal, P., Ployhart, R., & Marangoni, A. (2003). Learning by doing something else: Variation, relatedness, and the learning curve. *Management Science*, 49(1), 39–56.
- Schmoch, U. (1993). Tracing the knowledge transfer from science to technology as reflected in patent indicators. *Scientometrics*, 26(1), 193–211.
- Schmoch, U. (1997). Indicators and the relations between science and technology. *Scientometrics*, 38(1), 103–116.
- Schumpeter, J. A. (1939). *Business cycles: A theoretical, historical, and statistical analysis of the capitalist process*. New York: McGraw-Hill.
- Shane, S. (2004). *A general theory of entrepreneurship: The individual-opportunity nexus*. Cheltenham: Edward Elgar.
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265–269.
- Small, H., & Upham, P. (2009). Citation structure of an emerging research area on the verge of application. *Scientometrics*, 79(2), 365–375.
- St. John, C., & Harrison, J. S. (1999). Manufacturing-based relatedness, synergy, and coordination. *Strategic Management Journal*, 20, 129–145.
- Sternitzke, C. (2009). Patents and publications as sources of novel and inventive knowledge. *Scientometrics*, 79(3), 551–561.
- Swanson, D. R. (1986). Fish-oil, Raynaud's Syndrome, and undiscovered public knowledge. *Perspectives in Biology and Medicine*, 30(1), 7–18.
- Swanson, D. R. (1987). Two medical literatures that are logically but not bibliographically connected. *Journal of the American Society for Information Science*, 38(4), 228–233.
- Tamada, S., Naito, Y., Kodama, F., Gemba, K., & Suzuki, J. (2006). Significant difference of dependence upon scientific knowledge among different technologies. *Scientometrics*, 68(2), 289–302.
- Thomas, P. (2001). A relationship between technology indicators and stock market performance. *Scientometrics*, 51(1), 319–333.
- Thomke, S. (1998). Managing experimentation in the design of new products. *Management Science*, 44(6), 743–762.
- Tijssen, R. J. W., Buter, R. K., & Van Leeuwen, T. N. (2000). Technological relevance of science: An assessment of citation linkages between patents and research papers. *Scientometrics*, 47(2), 389–412.
- Tong, X., & Frame, J. D. (1994). Measuring national technological performance with patent claims data. *Research Policy*, 23, 133–141.
- Van Looy, B., Zimmermann, E., Veugelaers, R., Verbeek, A., Mello, J., & Debackere, K. (2003). Do science-technology interactions pay off when developing technology? An exploratory investigation of 10 science-intensive technology domains. *Scientometrics*, 57(3), 355–367.
- Verbeek, A., Debackere, K., & Luwel, M. (2003). Science cited in patents: A geographic “flow” analysis of bibliographic citation patterns in patents. *Scientometrics*, 58(2), 241–263.
- von Wartburg, I., Teichert, T., & Rost, K. (2005). Inventive progress measured by multi-stage patent citation analysis. *Research Policy*, 34(10), 1591–1607.
- Wang, S.-J. (2007). Factors to evaluate a patent in addition to citations. *Scientometrics*, 71(3), 509–522.
- Yetisgen-Yildiz, M., & Pratt, W. (2008). Evaluation of literature-based discovery systems. In P. Bruza & M. Weeber (Eds.), *Literature-based discovery* (pp. 101–113). Berlin: Springer-Verlag.