AWAY FROM THE STREETLIGHT: USING EXAMINER REJECTIONS TO CREATE MORE RELEVANT PATENT CITATION NETWORKS

Daniel R. Cahoy[*]

U.S. patent citations have been widely described in the literature as a means for linking one patent to another. Two types of citations can occur. A "backward citation" refers to a prior patent that was considered during the "citing" patent's examination. A "forward citation" refers to the consideration of a given patent in the future examination of another patent. Both are easily identified from existing patent databases. Researchers in law, economics, finance and management use these citations to characterize different aspects of the invention landscape.

Significantly, the utility of patent citation analysis is premised on the idea that a citation link indicates a meaningful connection between two patents. However, an assessment of actual patent citation behavior makes clear that this is frequently not the case. Many, if not most, patent citations have only a tangential relation to the citing patent, serving as mere background or alternate solutions to the same problem. While current techniques are useful, a more relevant citation link would constitute a significant advance for making any of assessments described above.

This research explores one very promising possibility for increasing citation relevance: the use of only those citations that served as the basis of a formal patentability rejection during the examination. Such citation links are relevant by definition, and are directed to the core of the patented subject matter. Until recently, the information necessary to identify this subgroup of highly relevant links was difficult to access and/or extremely expensive; thus, it has largely been ignored in the literature. New electronic means of accessing patent examination histories has opened the door for this analysis. This research employs government and commercial databases to explore this new measure and establish whether it offers a sufficient improvement over existing citation analysis.

WORKING DRAFT – NOT FOR CITATION OR PUBLICATION
Academy of Legal Studies in Business Annual Meeting
August 4-8, 2009
Denver, Colorado

---

[*] Associate Professor of Business Law, Smeal College of Business, the Pennsylvania State University. The author greatly appreciates comments on this research from participants at the Huber Hurst Annual Research Seminar, Boston, MA 2009, the Intellectual Property Scholars Conference, Des Moines, IA, Feb. 2008 and the Smeal College of Business Faculty Networking Lecture Series, Jan. 2008. In addition, the author wishes to thank Nicholas Fernez for his excellent research assistance and diligence in collecting data from most of the over 500 patent prosecution histories used in this article. Finally, the author wishes to acknowledge Stanford University for granting early access to data in the Stanford IP Clearinghouse. Funding for this research was provided by the Smeal College of Business Summer Research Grants Program.

Contents

Introduction 3
I. Characterizing Patent Ownership and Control 4
A. Patent Landscapes 6
B. Patent Neighborhoods 8
II. Problems in Empirical Assessment of Patent Environments 9
A. Ambiguity and Complexity in Qualitative Patent Assessment 10
B. Patent Citation Analysis as an Imperfect Substitute for Qualitative Assessment 12
1. References as a Proxy for Substantive Invention Linkage 15
2. The Mechanics of Patent Citation 17
3. The Limits of Current Patent Citation Analysis as an Analytical Tool 19
III. Increasing the Value of Citation Analysis by Utilizing Examination Information 20
A. Rejections Evince an Expert Opinion on Relevance 21
B. Empirical Evidence of Increased Relevance of 102/103 Citations 24
1. Validation with Density Based on Increased Incidence of Patent Pooling 25
2. Validation with Density based on Involvement in Litigation 27
3. Discussion of Study Results 30
Conclusion 31

INTRODUCTION

In the midst of the worst financial crisis in decades, perhaps no business is under more pressure than the U.S. automobile industry. Sales and profits have dropped precipitously,[1] and a major financial bailout may be necessary to sustain domestic automobile manufacturers in the United States. Many believe a long term solution to the woes of the industry lie in changing the business paradigm by investing in research and development and producing more innovative products.[2] In a related development, the nature of the fuel that powers those automobiles has faced increased scrutiny as policy makers become ever more concerned by the control held by foreign powers.[3] Again, the solution is viewed as increased research and development to produce alternative fuel sources.[4] In order for firms to make profitable investments in research and development, they must secure some return through legal instruments like patents.[5] In turn, they must avoid the patent rights already held by others and invent in an open space. In this context, as well as many, many others, the ability to ascertain opportunities for innovation by understanding the patent environment is critical.

Not surprisingly, the desirability of characterizing the patent environment has inspired a great deal of empirical investigation. For the most part, information on the linkages between multiple patents has been derived from the analysis of publicly available patent office database fields, particularly citations to other patents. It has the advantage of permitting one to construct networks of huge numbers of patents at once —potentially all of the patents issued since 1976[6] — and suggests a comprehensive

---

[1] Kate Linebaugh, *Auto Rescue Won't Jump Start Sales*, WALL ST. J. ONLINE (Dec. 19, 2008), http://online.wsj.com/article/SB122971236117722153.html ("Auto sales have fallen to their slowest pace since 1982, falling 37% in November from a year ago, according to Autodata Corp.").

[2] *See*, *e.g.*, Thomas L. Friedman, *How to Fix a Flat*, NY TIMES, Nov. 12, 2008, at A.31.

[3] *See* COUNCIL ON FOREIGN RELATIONS, NATIONAL SECURITY CONSEQUENCES OF U.S. OIL DEPENDENCY 22-23 (2006).

[4] *See*, *e.g.*, Stephan Herrera, *Bonkers about Biofuels*, 24 NATURE BIOTECHNOLOGY 755, 755 (2006).

[5] For example, in concert with the increasing interest in biofuels, there has been a significant rise in biofuel-related patents. *See* Daniel R. Cahoy & Leland Glenna, *Private Ordering and Public Energy Innovation*, 36 FLA. ST. L. REV. (forthcoming 2009) (depicting the rise in ethanol-based biofuel patenting over the last twenty years); Thomson-Reuters, Innovation Hot Spots 3 (2009), http://www.feedblitz.com/t.asp?/151161/320908/http://ip.thomsonreuters.com/media/pdfs/InnovationHotSpots_June2009.pdf (describing the huge growth in patents on algae-based biofuels).

[6] The publicly available database from the United States Patent and Trademark Office

understanding of the patent environment. To be sure, this information may provide a useful superficial overview of patent activity. Unfortunately, it is also quite noisy and far less relevant than necessary to make the substantive conclusions that many desire. Despite this detraction, the ease of access continues to attract researchers. In many respects, it is reminiscent of the famous story of the man looking for his keys under a streetlight.[7] A police officer approaches and discovers that the man actually lost his keys across the street. "Why are you searching under this streetlight if you lost your keys over there?," the police officer inquires. "Because the light is better here," the man answers.

This research strives to highlight the advantages of looking away from the metaphorical streetlight by delving deeper into harder-to-access patent documents in order to create more relevant linkages. More particularly, it suggests the use of citations related to a patent examiner's novelty and obviousness rejections as a proxy for related inventions. To detail this notion, the article begins in Part I by discussing the need to characterize the patent landscape to understand a particular innovation environment. In Part II, the article describes common empirical techniques for characterizing patent landscapes, focusing on citation analysis as the most widely used. It explains the shortcomings in current analytical techniques and the need for a better system. In Part III, the article proposes the use of a subset of patent citations in creating networks that are derived from patent examiner rejections. This section details the general theory and presents empirical data that validate the utility of this novel method of citation analysis. By using a more targeted form of citation analysis, the paper concludes, patent networks can be a more reliable tool for businesses and policy makers in innovation-related decision making.

I. CHARACTERIZING PATENT OWNERSHIP AND CONTROL

Patents provide their owners with the ability to exclude others from making, using, selling or offering to sell the claimed invention during the patent term.[8] From the perspective of an owner, patents are market exclusion devices that effectively provide protection for follow-on investment and prevent competitors from free-riding on the invention. From the perspective of a potential competitor, patents create a significant barrier to entry unless it is possible to design around the claims. If an

---

(PTO) contains full-text patents from 1976 to the present day. USPTO, *Patent Full-Text Database Contents*, http://patft.uspto.gov/help/contents.htm. Earlier patents dating from the early days of the U.S. government are available as image files. *Id.*

[7] *See*, *e.g.*, Robert M. Lawless & Stephen P. Ferris, *Professional Fees and Other Direct Costs in Chapter 7 Business Liquidations*, 75 WASH. U. L.Q. 1207, 1210 n. 10 (1997) (noting that the story is famous among empiricists).

[8] 35 U.S.C.§ 271(a) (2000).

invention proves to be essential in the market, and effective substitutes are not available, a patent can provide monopoly power.[9]

High technology firms have an obvious interest in patent rights. They may provide a cushion for product development that may take some time to catch on in the marketplace. For technology that would otherwise be viewed as easy to copy, this is often considered a critical advantage by investors and venture capitalists.[10] More recently, patent rights have been viewed as useful by a broader array of firms, as litigation defense devices and even marketing tools.[11] Interestingly, even in times of economic downturn when research and development spending is diminished, existing patent rights may be more important as firms attempt to squeeze what value they can out of their existing assets.[12]

The ever-greater reliance on patents has led to an increased need to assess patent environments. Before making an investment in research and development, a firm would be wise to know what rights already exist and what avenues may be foreclosed.[13] Future licensing opportunities may be identified and preliminary negotiations begun before it is too late for a firm to change course.[14] In addition, an investor should assess the patent environment to determine the degree of freedom a company has to grow and develop its products.[15] The potential that another company may rein in its progress through enveloping rights is a serious concern.

Beyond individual firms, there is a public policy rationale for assessing patent environments. In highly desired areas of technological innovation like alternative energy sources, nanotechnology or synthetic biology, the public has a significant interest in seeing the expansion of innovative activity.[16] Constraining intellectual property rights may provide

---

[9] *See* Edmund W. Kitch, *Elementary and Persistent Errors in the Economic Analysis of Intellectual Property Law*, 53 VAND. L. REV. 1727, 1730-31 (2000) (noting that whether patents provide monopoly power depends on the market).

[10] *See*, *e.g.*, James E. Malackowski & David I. Wakefield, *Venture Investment Grounded in Intellectual Capital*, *in* FROM IDEAS TO ASSETS: INVESTING WISELY IN INTELLECTUAL PROPERTY 160-62 (Bruce Berman ed., 2002).

[11] *See*, *e.g.*, Barry L. Grossman, *Patents as Advertising Tool? An Age-Old Technique*, NAT'L L.J., Oct. 27, 2003, at S6.

[12] *See* Lynne Marek, *Amid the Bad News, Firms are Hiring: IP Work hasn't Been A Victim of Crisis; Paten Protection a Priority*, NAT'L L.J. Nov. 17, 2008, at 10.

[13] Suzzane Harrison & Kevin G. Rivette, *The IP Portfolio as a Competitive Tool*, *in* PROFITING FROM INTELLECTUAL CAPITAL 123-24 (Patrick H. Sullivan, ed. 1998) (describing "cluster and bracket analysis" of competitor IP to get an idea of opportunities that may exist).

[14] *See* Kevin G. RIVETTE & DAVID KLINE, REBRANTS IN THE ATTIC, 113-116 (2000) (detailing the need to conduct a preliminary assessment of IP rights as part of an R&D strategy).

[15] *See* Malackowski & Wakefield, *supra* note __, at 160-62.

[16] *See generally* Sapna Kumar & Arti Rai, *Synthetic Biology: The Intellectual Property Puzzle*, 85 TEX. L. REV. 1745 (2007) (discussing the implications of patent ownership and

unexpected barriers necessitating a revision in intellectual property rules.[17] To ensure that the rules are revised in a manner that supports innovation and progress, it is essential that policymakers truly understand the current and future state of intellectual property ownership and control.

Depending on the need, a useful assessment of the patent environment can come at macro or micro level. Comparing the success of one country's patent regime to another may require broad, cross-technology data.[18] On the other hand, one may assess patents at the level of only one type of innovation within a larger technology classification.[19] The essential task is identifying the technology space and placing patent rights within that space to produce a landscape that reflect current and future invention and innovation potential.

A. Patent Landscapes

Many believe that control over important patent rights may pose a significant barrier to innovation. An example of such control would be the ownership of a patent covering a foundational invention, the use of which is so essential to an entire industry that it becomes a standard.[20] Another

---

control in the emerging field of synthetic biology); Mark A. Lemley, *Patenting Nanotechnology*, 58 STAN. L. REV. 601 (2005) (discussing the impact of patents on the nanotechnology innovation).

[17] *See* Peter Lee, *The Evolution of Intellectual Infrastructure*, 83 WASH. L. REV. 39, 93 (2008) (describing such patents as part of the "intellectual infrastructure" and arguing that "Absent efficient licensing, discussed below, exclusive rights on this infrastructure may inhibit myriad downstream applications."); Cahoy & Glenna, *supra* note 5, at __ (explaining how patent rights can serve as innovation barriers in some instances).

[18] *See* Walter G. Park and Juan Carlos Ginarte, *Intellectual Property Rights and Economic Growth*, CONTEMP. ECON. POL'Y, July 1997, at 51 (investigating the link between patent "strength" and R&D investment using data from over sixty countries from 1960 to 1990).

[19] For example, authors have attempted to assess the impact of patent activity related to so-called business methods, which are often relegated to a particular PTO classification (705). *See*, *e.g.*, John R. Allison & Emerson H. Tiller, *The Business Method Patent Myth*, 18 BERKELEY TECH. L.J. 987 (2003).

[20] *See* HERBERT HOVENKAMP ET AL., IP AND ANTITRUST: AN ANALYSIS OF ANTITRUST PRINCIPLES APPLIED TO INTELLECTUAL PROPERTY LAW § 35.1a (2002 & Supp. 2003); Philip B. Nelson, *Patent Pools: An Economic Assessment of Current Law and Policy*, 38 RUTGERS L.J. 539, 544-45 (2007) ("The establishment of industry standards, which often takes years to complete, can give market power to individual patents that was not present before the standard-setting process began."); Mark A. Lemley, *Ten Things to do about Patent Holdup of Standards (and One Not to)*, 48 B.C. L. REV. 149, 154-55 (2007) (describing how irreversible commitments to standards elevate the power a patent owner has to extract royalties); *See* Mark A. Lemley & David McGowan, *Legal Implications of Network Economic Effects*, 86 CAL. L. REV. 479, 523 (1998) ("To the extent intellectual property rights confer ownership interests in a strong network standard, they may create

example is a firm's consolidation of important patent rights into a portfolio that is difficult to design around.[21] In either case, the patent owner has little incentive to provide access as wide as the public would desire given that optimal monopoly profits are derived at a more constrained level.[22] In the context of new technology, it is possible that a powerful patent owner may even work to suppress a technology, particularly if it competes against one of the patent owner's existing products or services.[23] One of the more famous cases of supposed innovation-crushing consolidation involved Myriad Genetics' patent on the BRCA1 and BRCA2 breast cancer-linked genetic mutations.[24] According to many critics, Myriad's control over these essential pieces of the genome confer inordinate power, which Myriad allegedly uses to suppress research that would interfere with its business model.[25] Recently, Myriad's patents have been the subject of a long-shot lawsuit brought by the American Civil Liberties Union and PUBPAT claiming in part that enforcement of the patents violates the U.S. Constitution's First Amendment.[26]

But the monopoly case is fairly unusual. In practice, single patents rarely dominate technologies or provide anything close to market power. Rather, groupings of patents tend to dictate market dynamics. Patents over competing products may lock down a portion of the market.[27] More importantly, several patents may cover different aspects of the *same* product. This overlap occurs because patents provide no positive use rights – they confer only the right to exclude – meaning that a patentee is always at the mercy of another (or several others) with a blocking patent.[28] In addition, narrow patents may cover improvements of broader patents, further complicating the environment. In order to successfully market a product or utilize a process, a company may need to acquire use rights to

---

durable market power in network markets.")

[21] *See* Michael R. Ward, *Emerging Competition Policy Issues in Agricultural Biotechnology*, 44 AM. BEHAVIORAL SCIENTIST 504, 512-14 (2000) (describing the implications of consolidation in the field of agricultural biotechnology).

[22] *See* Lee, *supra* note __, at 93.

[23] *See* Kurt M. Saunders, *Patent Nonuse and the Role of Public Interest as a Deterrent to Technology Suppression*, 15 HARV. J.L. & TECH. 389, 402-17 (2002) (describing several cases of alleged patent suppression).

[24] U.S. Patent No. 5,693,473 (filed June 7, 1995).

[25] Richard Li-dar Wang, *Biomedical Upstream Patenting and Scientific Research: The Case for Compulsory Licenses Bearing Reach-Through Royalties*, 10 YALE J.L. & TECH. 251, 293-302 (2007).

[26] *See ACLU and PUBPAT Challenge Patents on Breast Cancer Genes* (May 12, 2009), http://www.pubpat.org/brcafiled.htm.

[27] *See*, *e.g.*, Daniel R. Cahoy, *Confronting Myths and Myopia on the Road from Doha*, 42 GA. L. REV. 131, 144-45 (2007) (describing the oligopolistic behavior of drug companies when new classes of compounds like statins are introduced).

[28] *See* Christopher M. Holman, *Impact of Human Gene Patents on Innovation and Access: A Survey of Human Gene Patent Litigation*, 76 UMKC L. REV. 295, 302 (2007).

dozens or even hundreds of patents.[29]

In recent years, the problem of overlapping patents has received greater attention than the monopoly question. The term "anticommons" was popularized by the research of Rebecca Eisenberg and Michael Heller, which suggested that too many conflicting property rights could result in the underutilization of a technology field.[30] A related situation in which an innovator cannot easily determine who owns important rights and is not able to commercialize has been termed a patent "thicket."[31] Both situations are widely accepted and commented on in academic circles, though empiricists have experienced difficulty in identifying the existence of thickets beyond anecdotal cases.[32]

B. Patent Neighborhoods

Although consolidation and dispersion problems could theoretically impact any technology area, there appears to be significant differences between industries. The coverage of a technology area and degree of patent overlap, otherwise known as patent density, seems to depend on various technology centric factors.[33] For example, some technologies are known as "crowded arts" due to the huge number of patents and space for only incremental innovation.[34] In some technically complex crowded arts, like the semiconductor industry, a number of invention rights may read on a single product.[35] Thickets may be more likely to develop in these

---

[29] *See* Mark A. Lemley, *Are Universities Patent Trolls?*, 18 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 611, 614-19 (2008)

[30] Michael Heller & Rebecca Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCI. 698, 698-99 (1998).

[31] Carl Shapiro, *Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting*, *in* 1 INNOVATION POLICY AND THE ECONOMY 119, 124-26 (Adam B. Jaffe et al. eds., 2001) (describing as a basis of patent thicket, "holdup" problem when "hundreds if not thousands of patents... can potentially read on [the same] product").

[32] *See* Gavin Clarkson, *Patent Informatics for Patent Thicket Detection: A Network analytic approach for measuring the density of patent space*, at 1-2 (2005) (presented at the Academy of Management Conference, Honolulu), *available at* http://w4.stern.nyu.edu/emplibrary/ACFNpuvR7.pdf.

[33] Dan L. Burk & Mark A. Lemley, *Policy Levers in Patent Law*, 89 VA. L. REV. 1575, 1629-30 (2003) (noting that a single economic theory of bargaining does not fit all technologies because patent use and scope can be quite different).

[34] *See* Augustine Med., Inc. v. Gaymar Indus., Inc., 181 F.3d 1291, 1301-02 (Fed. Cir. 1999) ("Without extensive prior art to confine and cabin their claims, pioneers acquire broader claims than non-pioneers who must craft narrow claims to evade the strictures of a crowded art field. Thus, claim scope itself generally supplies broader exclusive entitlements to the pioneer.");

[35] Lemley, *supra* note 23, at 613; Doug Lichtman & Mark A. Lemley, *Rethinking Patent Law's Presumption of Validity*, 60 STAN. L. REV. 45, 55 (2007); FEDERAL TRADE COMMISSION (FTC), TO PROMOTE INNOVATION: THE PROPER BALANCE OF COMPETITION AND PATENT LAW AND POLICY 34

environments.[36] Conversely, there is relatively little patent activity in arts like "leather manufactures" (classification 069), and one would be surprised to find a great deal of overlap and conflict there.[37]

The result of industry-specific differences is that one must conduct independent assessments of patent ownership to determine whether invention and innovation-suppressing ownership patterns exist. What is true for the semiconductor industry is not necessarily true for pharmaceuticals or agricultural biotechnology.[38] This relatively obvious point is often obscured by the fact that most patent systems are unitary in structure.[39] In other words, from a public policy perspective, patent rights are treated equivalently regardless of the type of technology (with inventions related to software and business methods among the few exceptions). Therefore, it is somewhat of a challenge to consider patent environments specifically by technology type.

II. PROBLEMS IN EMPIRICAL ASSESSMENT OF PATENT ENVIRONMENTS

Patents truly lend themselves to empirical analysis, providing useful information on potential market power and the competitive nature of innovation environments. Not only are they discrete property rights with statutorily defined powers, but also it is easy to obtain essential information. Patents are public information, and are available from government databases without charge.[40] Several private databases also provide patent information in more easily searchable formats or with additional databases included.[41] It is possible to search for information within several well-defined fields, such as inventor or assignee name, technology classification, claim language, and related references.[42] This permits detailed analysis on

---

[36] *See* Burk & Lemley, *supra* note __, at 1627-28 (noting that the thicket paradigm maps nicely to the semiconductor field).

[37] According to the USPTO's patent statistics, technology class 069 had only 52 patents issue from 1997-2006. USPTO, *Patent Counts by Class by Year: January 1977-December 2006*, http://www.uspto.gov/web/offices/ac/ido/oeip/taf/cbcby.htm (last modified Oct. 26, 2007). This is in great contrast to class 515, "Drug, Bio-Affecting and Body Treating Compositions," in which 30,175 patents issued over the same period.

[38] *See* Burk & Lemley, *supra* note __, at 1629-30.

[39] *See* NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., A PATENT SYSTEM FOR THE 21ST CENTURY 83-84 (Stephen A. Merrill et al. eds., 2004) (describing the historical reticence to depart from a unitary patent system, and generally supporting the notion that the unitary nature of the U.S. system be kept intact).

[40] One the U.S. Patent & Trademark Office's website, one may search issued patents and pending applications. USPTO, *Patent Full-Text and Image Databases*, http://patft.uspto.gov/.

[41] For example, Delphion permits one to search U.S., European Patent Office, Japanese and German patents at the same time. *See* http://www.delphion.com/research/.

[42] USPTO, *Patent Full-Text and Image Database*,

the characteristics of patent ownership and the transfer of knowledge across countries.

It naturally follows that an extensive literature has evolved related to the empirical assessment of patent environments. Essentially, researchers describe three steps: (1) determining the patents that are relevant to the technology; (2) identifying the owners of those patents; and (3) determining which patents overlap in the same innovation space. Empirical assessments of the first two points have existed for decades.[43] The third step is of more recent interest and is generally perceived to be much more difficult, resulting in the use of creative, indirect measures.

A. Ambiguity and Complexity in Qualitative Patent Assessment

Although easy to obtain, the information from patent databases is far from succinct and clear in all respects. Most problematic is determining the scope and content of the rights conveyed in a patent. Patent rights are defined by their claims, much the same as a metes and bounds description in a real estate deed.[44] However, unlike a deed, much more interpretation is necessary to understand the meaning of a particular patent claim because patent rights do not occupy a physical space or embody particular objects.[45] The rights could cover a single characteristic of a huge variety of objects, depending on how the language is interpreted.[46] Claims are language-

---

http://patft.uspto.gov/netahtml/PTO/search-adv.htm (last viewed Jan. 7, 2009) (advanced search page describing thirty-one different database fields).

[43] *See* Bronwyn H. Hall et al., *The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools*, *reprinted in* PATENTS, CITATIONS, AND INNOVATIONS: A WINDOW ON THE KNOWLEDGE ECONOMY 405-07 (Adam Jaffe & Manuel Trajtenberg eds., 2002).

[44] *See, e.g.,* Corning Glass Works v. Sumitomo Elec. U.S.A., Inc., 868 F.2d 1251, 1257 (Fed. Cir. 1989) ("A claim in a patent provides the metes and bounds of the right which the patent confers on the patentee to exclude others from making, using or selling the protected invention.").

[45] *See* Pfaff v. Wells Elecs., Inc., 525 U.S. 55, 60-61 (1998) ("The primary meaning of the word 'invention' in the Patent Act unquestionably refers to the inventor's conception rather than to a physical embodiment of that idea. The statute does not contain any express requirement that an invention must be reduced to practice before it can be patented."). This is true of all intellectual property — the rights are separable from any embodiment. See Susan Scafidi, *Intellectual Property and Cultural Products*, 81 B.U. L. REV. 793, 799-800 (2001). Indeed, no physical embodiment need exist for one to qualify for a patent. *See* Sean B. Seymore, *Heightened Enablement in the Unpredictable Arts*, 56 UCLA L. REV. 127, 143(2008) ("In contrast to the canons for scientific research and publishing, an inventor can obtain a patent without conducting a single experiment . . . It is well settled in U.S. patent law that conception, and not any physical act, is the 'touchstone' of inventorship.") (citing Burroughs Wellcome Co. v. Barr Lab., Inc., 40 F.3d 1223, 1227-28 (Fed. Cir. 1994)).

[46] *See* Kumar & Rai,, *supra* note __, at 1751 (discussing the power of foundational

based, naturally subject to varying interpretation. Ultimately, courts provide the final guidance,[47] but until a litigation event occurs, one can only use a reasonable best assessment. Arguably, this is one of the limitations in using a language-based property system to set forth the borders of an invention, something that is by definition so new that conventionally descriptive terms may not exist. The consequence of patent claim ambiguity is that it can be very difficult to determine precisely what a patent covers.

Making this system even more complex, the subject matter covered by patent claims can change over time depending on the decisions of the courts (both related to the patent in question, and patent law concepts in general). Patent law is an extremely fluid field, and cases that narrow the reach of claim language[48] or even eliminate patentable subject matter[49] *ex post* to the prosecution of a patent are not uncommon. When such decisions are reached by the courts, the revision of rights applies retroactively, impacting settled patent rights.[50]

While property rights ambiguities also apply to all other intellectual

---

patents that impact multiple embodiments that may be produced by several different firms).

[47] In the United States, claim interpretation is a question of law, which means that a court will decide the boundaries (outside of the Patent Office). Markman v. Westview Instruments, Inc., 517 U.S. 370, 390-91 (1996); Kemco Sales, Inc. v. Control Papers Co., 208 F.3d 1352, 1359-60 (Fed. Cir. 2000). Within a litigation, this interpretation is usually conducted during a so-called *Markman* hearing before the trial takes place. James Ware, *Patent Rules of Evidence*, 23 SANTA CLARA COMPUTER & HIGH TECH. L.J. 749, 752 (2007). A particular district court's interpretation is not binding on any other court, thus a patent may be subject to multiple *Markman* hearings over its lifetime. *See*, *e.g.*, Wilson Sporting Goods Co. v. Hillerich & Bradsby Co, 2003 WL 21911241 at *3 (N.D. Ill. Aug. 8, 2003), *vacated on other grounds*, 442 F.3d 1322 (Fed. Cir. 2006). Perhaps due to the inherent uncertainty of claim language, the courts have developed a rule called the "doctrine of equivalents" for supplementing the articulated boundaries when it is apparent that the invention is broader than claimed. Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co., 535 U.S. 722, 731 (2002) ("Unfortunately, the nature of language makes it impossible to capture the essence of a thing in a patent application."). *But see* John R. Allison & Mark A. Lemley, *The (Unnoticed) Demise of the Doctrine of Equivalents*, 59 STAN. L. REV. 955, 956-57 (2007) (noting that the impact of the doctrine appears to be in decline).

[48] *See*, *e.g.*, NTP, Inc. v. Research in Motion, Ltd, 418 F.3d 1282, 1317-18 (Fed. Cir. 2005) (determining for the first time that a method must be entirely practiced in the U.S. to infringe a patent).

[49] *See*, *e.g.*, In re Bilski, 545 F.3d 943, 959-60 (Fed. Cir. 2008) (deciding that a prior, broader test for patentability was inapplicable and substituting it with a narrower one, potentially eliminating existing patent rights), *cert. granted*, Bilsky v. Doll, __U.S.__, 2009 WL 221232 (June 1, 2009).

[50] Daniel R. Cahoy, *Changing the Rules in the Middle of the Game: How the Prospective Application of Judicial Decisions Related to Intellectual Property Can Promote Economic Efficiency*, 41 AM. BUS. L.J. 1, 29-38 (2003) (describing the current state of the law of retroactivity and finding that no restriction currently exists on court decisions that eliminate property rights).

property types, the situation is arguably more problematic with patents due to the likelihood of overlap between rights, as described above. The need for a more efficient mechanism for assessing patent environments is apparent. And one should be willing to sacrifice some degree of accuracy if it means that large scale assessment is reasonably possible.

B. Patent Citation Analysis as an Imperfect Substitute for Qualitative Assessment

Instead of embarking on the extraordinarily time-consuming and often ambiguous task of reviewing the claims of individual patents, researchers have long searched for some more abstract and easily quantifiable indicia of the patent environment's density. The simplest approach is to treat all patents as equal allocations of rights and tally their numbers.[51] One could argue that the number of patents within a technology class or owned by a particular individual says something about how developed a particular technology area is.[52] Many academic works undertake such an analysis,[53] but the limitations are fairly obvious. Because patents can have such varying scope, and it is difficult to ensure that any artificial grouping of patents actually involves substantive connections, patent counting can lead to exaggerated conclusions.[54]

To move beyond patent counting, a variety of patent metrics are available, and researchers have used them to measure attributes like density[55], value[56] and knowledge transfer (*see* Table 1).[57]

---

[51] *See*, *e.g.*, JOHN L. KING, CONCENTRATION AND TECHNOLOGY IN AGRICULTURAL INPUT INDUSTRIES 9-11 (2001), http://www.ers.usda.gov/publications/aib763/aib763.pdf (assessing consolidation in agricultural biotechnology by counting the number of patents owned by particular firms).

[52] *See* Zvi Griliches, *Patent Statistics as Economic Indicators: A Survey*, 28 J. ECON. LIT. 1661, 1701-02 (1990) ("In spite of all of the difficulties, patent statistics remain a unique source for the analysis of the process of technical change.").

[53] Hall et al, *supra* note __, at 405-07.

[54] For example, a firm may file multiple patents on different embodiments of the same basic invention. Double patenting concerns are addressed by the use of "terminal disclaimers." USPTO, MANUAL OF PATENT EXAMINING PROCEDURE (MPEP) § 804.02 (8th ed., rev. 7, 2008). In such a case, the scope of several patents may not be as large as one broad patent. Yet simple patent counting would associate more market power with the greater number of patents.

[55] Patent density is a representation of both the number of patents in a technological space as well as the extent to which they are connected. Patent-dense fields are difficult to navigate without infringing on another's claims. Extremely dense fields may be characterized as a thicket. *See supra* notes 28-29, and accompanying text.

[56] The value of a patent can be superficially characterized as its ability to provide an economic advantage to its owner, such as additional income, relief from competitor royalties, reducing competition, etc. Although there are many alternatives for calculating patent value, a metric may be effective by providing only a general indication.

[57] Many studies have considered how knowledge from one source may be reflected in the patent of another. *See infra* note 83. Using connections to patents such as citations is a

means of tracking this transfer of information, providing information on invention influences.

Table 1
Examples of Metrics Used in Empirical Patent Analysis[58]

Patent Metric	Indicates Density	Indicates Value	Tracks Knowledge Transfer
Patent Counting[59]	X	X	X
Backward Citation[60]	X		X
Forward Citation[61]	X	X	X
Renewal Data[62]		X	
Involved in Litigation/Reexamination[63]	X	X	
Claim Number[64]		X	
Specification Length	X	X	
Neural Net Connections[65]	X		X
Existence of Patent Pool[66]	X	X	X

[58] Allison, et al., cite additional methods in greater detail. Allison et al., *Valuable Patents*, 92 GEO. L.J. 435, 439-43 (2004). Many articles employ combined indices that utilize multiple metrics. Dietmar Harhoff, *et al*., *Citations, Family Size, Opposition and the Value of Patent Rights*, 32 RESEARCH POL. 1343, 1343-44 (2003) (describing new study that correlates various patent characteristics with patent quality); Jean O. Lanjouw & Mark Schankerman, *Patent Quality and Research Productivity: Measuring Innovation With Multiple Indicators*, 114 ECON. J. 441, 442 (2004) (study that develops "a composite index of patent quality using multiple characteristics of patents).

[59] *See*, *e.g.*, KING, *supra* note __.

[60] John R. Allison & Emerson H. Tiller, *The Business Method Patent Myth*, 18 BERKELEY TECH. L.J. 987, 1082 (2003); Dietmar Harhoff & Markus Reitzig. *Determinants of Opposition Against EPO Patent Grants--The Case of Biotechnology and Pharmaceuticals* (Working Paper, May 2000)

[61] *See*, *e.g.* Brownwyn H. Hall, et al, *Market Value and Patent Citations*, 36 RAND J. ECON. 16, 16-17 (2005).

[62] *See*, *e.g*, Mark Schankerman, *How Valuable is Patent Protection? Estimates by Technology Field*, 29 RAND J. ECON. 77, 78-79 (1998); Mark Schankerman & Ariel Pakes, *Estimates of the Value of Patent Rights in European Countries During the Post-1950 Period*, 96 ECON J. 1052 (1986) (weighting patent value by analyzing renewal data); Moore, *supra* note __, at 1530.

[63] Jean O. Lanjouw & Mark Schankerman, *Characteristics of Patent Litigation: A Window on Competition*, 32 RAND J. ECON. 129 (2001); Allison et al. *supra* note __; John Allison & Thomas W. Sager, *Valuable Patents Redux On the Enduring Merit of Using Patent Characteristics to Identify Valuable Patents*, 85 TEX. L. REV. 1769 (2007).

[64] *See*, *e.g.* Jean O. Lanjouw & Mark Schankerman, *Stylized Facts of Patent Litigation: Value, Scope and Ownership*, at 17-18 (Nat'l Bureau of Econ. Research, Working Paper No. 6297, 1997).

[65] Clarkson, *supra* note __.

[66] *Id.*

Triadic Global Patenting[67]	X	X	X
Size of Owning Entity[68]	X	X	X

In the specific context of connections between patents, particularly useful information is derived from references to prior patents. These citations are believed by many to provide useful information on technologies with a substantive connection.[69] In addition, the number of citations may provide helpful information. It is useful to understand the rational for the use of citations and the particulars for how citations appear. From there, the problems can be appreciated as well as the potential for a better system.

1. References as a Proxy for Substantive Invention Linkage

A reference to another work generally provides some indication that the referenced material was important in producing the new work. This principle is a generally accepted tenet of citation analysis that can be applied in a number of different formats. One of the most common formats is academic articles.[70] The citation of another's work suggests at least a connection in subject matter, and perhaps a great influence on the analysis or even conclusions.[71] More citations would be considered as evidence that the work has ties to a broader array of other works.[72] Moreover, the "forward citation" of a work by subsequent authors is usually considered to be a very strong measure of significance or impact.[73]

---

[67] The Organization for Economic Cooperation and Development's (OECD) Patent Project, which is conducted by the U.S. National Science Foundation (NSF), the World Intellectual Property Organization (WIPO), the European Union, and the patent offices of the U.S., Japan and the EU, uses a "triadic" system that recognizes inventions that are covered by patents in the United States, Europe, *and* Japan over a given time period. *See* ORG. FOR ECON. COOPERATION AND DEV., COMPENDIUM OF PATENT STATISTICS (2005) [hereinafter OECD PATENT STAT.], *available at* http://www.oecd.org/dataoecd/60/24/8208325.pdf.

[68] Allison et al. *supra* note __.

[69] Manuel Trajtenberg,, A *Penny for your Quotes: Patent Citations and the Value of Innovations*, 21 RAND J. ECON. 172, 173 (1990).

[70] *See*, *e.g.*, Michael J. Stringer et al., *Effectiveness of Journal Ranking Schemes as a Tool for Locating Information*, 3 PLOS ONE e1683, e1683 (2008) ("[T]he entire science and technology community is relying more and more on citation-based statistics as a stool for evaluating the research quality of individuals and institutions.").

[71] *See*, *e.g.*, Jeppe Nicolaisen, *Citation Analysis*, 41 ANN. REV. INFO. SCI. & TECH. 609, (2007) (the list of basic assumptions in citation analysis includes the fact that "A cited document is related in content to the citing document.") (citing L.C. Smith, *Citation Analysis*, 30 LIBRARY TRENDS 83 (1981)).

[72] *Id.*

[73] *See* Stringer et al., *supra* note __, at e1683-84. Note that informal tools like Google scholar offer citation tools.

The same rationale for the use of citations in assessing academic works applies to patents, which, after all, are just literary descriptions of an invention.[74] The patent document itself usually contains references to prior inventions (as well as non-patent documents) that provide background and context.[75] In addition, the back and forth discussion between the patent examiner and the patentee also frequently involves the citation of patents as a means for determining the proper scope of the invention.[76] It is entirely reasonable to presume that these citations indicate a connection and suggest significance in exactly the same way as citations in an academic paper.

Although citation analysis is not a complete substitute for detailed claim comparison, one can see how it can easily provide useful, superficial information on connections between patented subject matter. Most importantly, large numbers of patents can be assessed in this manner.[77] Commercial databases even graphically depict the connections.[78] Thus, patent citation analysis has become one of the most common means of depicting patent landscapes.[79] Researchers in law, economics, finance and management use these citations to characterize different aspects of the invention landscape. For example, Gavin Clarkson utilized a network analytic technique to determine patent citation density and correlate thickets with patent pools.[80] Manuel Trajtenberg, Adam Jaffe and Bronwyn Hall used patent citations as a proxy for patent value.[81] Similarly, Robert Merges has employed citations as a measure of patent quality.[82] Adam

---

[74] Patent applications must include a "written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains . . . ." 35 U.S.C. § 112 (2000).

[75] Trajtenberg, *supra* note __, at 27; MPEP § 706.05 (8th ed, rev. 7, 2008) ("Allowed applications should generally contain a citation of pertinent prior art for printing in the patent, even if no claim presented during the prosecution was considered unpatentable over such prior art.").

[76] MPEP § 707.05(d) (8th ed., rev. 7, 2008).

[77] One of the most important advances in patent citation analysis came with the development of the NBER patent citation data file by Hall et al. in 2001. *See generally* Hall et al., *supra* note __ (describing the characteristics of the database) It contained information from 2,923,922 patents in searchable fields. *Id.* at 407. Today, one can purchase bulk patent data from the USPTO or simply use the website interface. *See* USPTO, *Important Notices Concerning the Patent Full-Text and Full Page Image Databases*, http://www.uspto.gov/patft/help/notices.htm (last modified April 6, 2006).

[78] *See*, *e.g.*, Delphion, *Using Citation Link*, http://www.delphion.com/help/citelink_help (last updated Jul. 20, 2004).

[79] *See* Juan Alcácer et al., *Applicant and Examiner Citations in US Patents: An Overview and Analysis*, Harv. Bus. School Working Paper 09-016, at p. 2 (2008) ("The empirical literature on technology and innovation increasingly relies on measures based on patent citations.").

[80] Clarkson, *supra* note __.

[81] Bronwyn H Hall, *Market Value and Patent Citations*, 36 RAND J. ECON. 16 (2005).

[82] Robert P Merges, Patents, *Entry and Growth in the Software Industry*, Working

Jaffe and others have employed patent citation data to track R&D and knowledge transfer.[83] Recently, John Allison and Thomas Sager validated several of these measures using a database of litigated patents.[84]

2. The Mechanics of Patent Citation

While it seems like a fairly fundamental question, the relevant literature does not generally address in much detail how citations actually end up in patent documents.[85] Though there are concrete rules, there is also room for individual decision-making. Thus, understanding the process is essential to determining the relevancy of the citation data.

In short, patent citations (as well as references to other relevant literature) can come from either the applicant or the patent examiner. When an applicant files a patent application, he or she is under a continuing duty to disclose information that is "material to patentability."[86] Generally, one accomplishes this by submitting an Information Disclosure Statement (IDS) that lists relevant references.[87] The number of references that can be submitted is unlimited,[88] and patent applicants and their attorneys are more likely to be over-inclusive than under. This is because the consequence for intentionally failing to disclose a material piece of information is the perpetual unenforceability of the patent, notwithstanding any argument that a patentable invention actually existed in the application.[89]

To the list of references submitted by the patent applicant (if any), the patent examiner may add additional citations.[90] These references are primarily uncovered as a result of the patent examiner's own search of the

---

Paper (2006) *available at* SSRN: http://ssrn.com/abstract=926204.

[83] *See*, *e.g.*, Adam Jaffe et al., *Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations*, 108 Q. J. ECON. 577 (2003); Adam Jaffe et al., Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors." 90 AM. ECON. REV. 215 (1990); F.M. Scherer, *Using Linked Patent and R&D Data to Measure Interindustry Technology Flows*, *in* R&D, PATENTS, AND PRODUCTIVITY *(* Z. Griliches ed., 1984); Xin Li, *Patent Citation Network in Nanotechnology (1976-2004)*, 9 J. NANOPARTICLE Res. 337 (2007).

[84] Allison & Sager, *supra* note __.

[85] The working paper by Alcácer et al. is a notable exception, containing considerable detail on the citation process and incentives. Alcácer et al., *supra* note __, at 5-10.

[86] 37 C.F.R. § 1.56 (2007) ("Each individual associated with the filing and prosecution of a patent application has a duty of candor and good faith in dealing with the Office, which includes a duty to disclose to the Office all information known to that individual to be material to patentability . . ..")

[87] 37 C.F.R. §§ 1.97, 1.98 (2007); MPEP § 609 (8th ed., rev. 7, 2008).

[88] MPEP § 609 (8th ed., rev. 7, 2008).

[89] *See* Digital Control Inc. v. Charles Mach. Works, 437 F.3d 1309, 1315-16 (Fed. Cir. 2006) (describing inequitable conduct based on failure to disclose a reference as type of fraud).

[90] MPEP § 707.05 (8th ed., rev. 7, 2008); Alcácer et al., *supra* note __, at 8-9.

art related to the invention.[91] The examiner does not duplicate the citations already made by the applicant,[92] and it is likely that much of what results from the search is already in the file.[93] In addition to references uncovered by a search, patent examiners experienced in working with a particular art may keep a file of references that have been cited in similar applications.[94] For the most part, this collection of references is the working group that is applied during the examination. Additional relevant references uncovered while the examination is ongoing may be included,[95] though these will be few in number.

Until recently, the *ex parte* nature of the exam[96] precluded third parties from possessing the knowledge necessary to contribute relevant references. But with the eighteen-month publication procedure that applies to most U.S. applications,[97] third parties could theoretically now provide useful references. However, the PTO only permits third parties to submit prior art for two months following publication.[98] Moreover, the examiner is not required to consider or cite any such submission.[99] The PTO has engaged in pilot programs that permit outside submission of prior art, but it has yet to adopt the programs on a wide scale.[100]

If the patent application proceeds to issuance (likely with many changes), the patent examiner identifies "pertinent" references for citation on the face of the patent.[101] Although this could theoretically be an opportunity for highlighting very close prior art, the pertinent limitation is generally applied liberally. Often, all of the art cited by the applicant and the examiner is deemed pertinent enough to appear on the face of the patent. If references are omitted, the examiner is under no obligation to offer an

---

[91] Alcácer et al., *supra* note __, at 8-9 (noting that the search can be quite idiosyncratic).

[92] MPEP § 1302.12 (8th ed., rev. 7, 2008) ("The examiner does not list references which were previously cited by the applicant (and initialed by an examiner) on an Information Disclosure Statement . . . .").

[93] Alcácer et al., *supra* note __, at 3 (stating that, on average, patent examiners add only about forty percent of patent citations to the file).

[94] There can be significant variation in an examiner's diligence in finding prior art. This variation appears to be related to experience and it can be correlated to the likelihood of obtaining a favorable disposition from that examiner. *See* Mark Lemley & Bhaven N. Sampat, *Eaminer Characteristics and the Patent Grant Rate*, 11-12 (Jan. 2009), *available at* http://ssrn.com/absract=1329091.

[95] In fact, if discovered by the patent applicant, disclosing them to the patent examiner is a duty. 37 C.F.R. § 1.56 (duty to disclose information material to patentability).

[96] 35 U.S.C. § 122(a) (2000) (confidentiality of applications).

[97] 35 U.S.C. § 122(b) (2000) (publication of applications after eighteen months).

[98] 37 C.F.R. 1.99 (2007)

[99] MPEP § 1134.01 (8th ed., rev. 7, 2008).

[100] *See* USPTO, *Peer-Reviewed Prior Art Pilot*, http://www.uspto.gov/web/patents/peerpriorartpilot/ (last visited Jan. 7, 2009).

[101] MPEP § 707.05 (8th ed., rev. 7, 2008).

explanation, and there typically is none.

3. The Limits of Current Patent Citation Analysis as an Analytical Tool

Given the liberal construction of the limitation for pertinence, patent citations – whether submitted by the applicant or added by the examiner – can have significance that varies dramatically. A reference may be critical to the patentability of an invention, forming the basis of a claim rejection for lack of novelty (known as a § 102 rejection)[102] or nonobviousness (a § 103 rejection).[103] On the other hand, a reference may simply describe an alternate method of accomplishing the invention, cited as background.[104] Both references are relevant, but the former is obviously much more important.

In view of the limitations of total citation analysis, one can see how citation networks can be misleading. Loosely connected patents may comprise a seeming web of impenetrable rights that in reality does not exist — a phantom thicket, as it were. The question of citation link reliability was addressed in a recent survey of inventors by Jaffe, et al.[105] In the survey, named inventors on issued patents were asked to rate the relatedness of patents cited in their own patent. Strikingly, inventors rated approximately 44% of cited patents as having a very low level of similarity of technology and application with the claimed invention.[106] The authors concluded "a significant fraction of citations are to patents judged by the inventors themselves to be unrelated, even if the judgment is made by a cited inventor."[107] This result is not surprising in view of the process for adding citations to a patent. Such results cast doubt on the reliability of constructed networks. A recent article accurately summed up the issues with current citation analysis:

> The mere existence of a citation from one patent to another cannot tell us [whether the patents overlap or are substitutes]. Determining

---

[102] 35 U.S.C. § 102 (2000) (patents must be novel with respect to the prior art).

[103] 35 U.S.C. § 103 (2000) (patents must be non-obvious advances over the prior art).

[104] *See*, *e.g.*, James Bessen & Michael J. Meurer, *Lessons for Patent Policy from Empirical Research on Patent Litigation*, 9 LEWIS & CLARK L. REV. 1, 23 (2005) ("Subsequent patents might cite earlier patents simply to distinguish them, and not because they build on them.").

[105] Adam B. Jaffe et al., *The Meaning of Patent Citations: Report of the NBER/Case-Western Reserve Survey of Patentees*, *reprinted in* PATENTS, CITATIONS, AND INNOVATIONS: A WINDOW ON THE KNOWLEDGE ECONOMY 379 (Adam Jaffe & Manuel Trajtenberg eds., 2002).

[106] *Id.* at 386, fig. 1a.

[107] *Id.* at 387.

> the meaning of a particular citation is an extremely labor-intensive process requiring understanding of the legal and technical relationship between the citing and cited patents.[108]

Yet empiricists continue to employ this analysis due to its relative ease, the fact that it has *some* relevance for making general assertions in the context of very large-scale analysis,[109] and perhaps a lack of understanding regarding the methodological detractions.

Clearly, the use of citations to construct networks of patents would be greatly assisted by a means of culling out the unrelated patents. One intriguing possibility is provided by Alcácer et al.'s work, which separates examiner-cited patents from the overall citations.[110] Such citations are marked on the face of the issued patent with an asterisk.[111] If one assumed that examiners only add highly relevant patents, this would seem to be a relatively good way of crafting a useful subgroup of citations. However, examiner-cited patents are not necessarily relevant to the patentability of an invention, as they may be added for other reasons (though it is reasonable to assume this occurs more rarely than in the general citation group). More importantly, examiner-cited patents only add to the population of references already cited by the applicant, never duplicating it.[112] Thus, if the applicant submitted the most related patent —i.e., the one that compels the most scrutiny for the claimed invention — it would be entirely missed by a system that analyzes only examiner-cited references.

Despite its limited applicability in this particular context, Alcácer et al.'s work highlights the advantages in using the examiner's expertise to identify truly important citations. A solution based on enhanced citation weighting has potential. Intriguingly, hope may lie in newly available patent information.

III. INCREASING THE VALUE OF CITATION ANALYSIS BY UTILIZING EXAMINATION INFORMATION

As described above, the limitations of current citation analysis exist

---

[108] Katherine J. Strandburg et al, *Law and the Science of Networks: An Overview and an Application to the "Patent Explosion,"* 21 BERKELEY TECH. L.J. 1293, 1347 (2006) (explaining that, while citation analysis is useful, it does not always identify thickets).

[109] *See* Kimberley A Moore, *Worthless Patents*, 20 BERKELEY TECH. L.J. 1521, 1532 n. 44 (2005) ("The citation of a large number of U.S. patent prior art references means that the invention is likely not a pioneering invention, but rather an improvement in a crowded field.").

[110] Alcácer, *supra* note __.

[111] *Id.* at 2-3; MPEP § 1302.12 (8th ed., rev. 7, 2008).

[112] MPEP § 1302.12 (8th ed., rev. 7, 2008)

because the most widely used databases contain information from the issued patent document only. It is necessarily a summary document, much like a real estate deed.[113] However, as patent practitioners are well aware, there is a great deal of useful information that is contained in the deeper, documentary history of the patent prosecution.[114] Known as the "prosecution history," "file history" or "file wrapper," it contains documents reflecting every single piece of paper exchanged between the applicant and the patent examiner.[115] Within this history is important insight into which references the patent examiner considers to be most important and connected. Accessing this assessment to help inform citation networks would be extremely useful, and it is now suddenly more available.

A. Rejections Evince an Expert Opinion on Relevance

The patent examination process in the United States (as well as other regions with rigorous examination criteria[116]) is essentially a back and forth negotiation on the patentability of the invention at issue.[117] In no way is prosecution a simple rubber stamp. Instead, it involves the attempt by the applicant to claim as much as possible against the resistance of the examiner to allow a patent on an invention that is already claimed by another or in the public domain.[118] Through this interaction, details about the claimed subject matter and the proximity of existing inventions are highlighted.

An examiner's primary bases for scrutinizing an applicant's invention in view of existing art are the dictates that patent claims be "novel"[119] and "non-obvious."[120] The novelty requirement, contained in 35

---

[113] *See, e.g.,* Corning Glass Works v. Sumitomo Elec. U.S.A., Inc., 868 F.2d 1251, 1257 (Fed. Cir. 1989).

[114] In fact, the documentary history is actually considered part of the patent. *See* Phillips v. AWH Corp., 415 F.3d 1303, 1317 (Fed. Cir. 2005) ("The prosecution history, which we have designated as part of the 'intrinsic evidence,' consists of the complete record of the proceedings before the PTO and includes the prior art cited during the examination of the patent.).

[115] *Id.*

[116] Many consider the Japanese Patent Office (JPO) and the European Patent Office (EPO) to have standards at least as stringent as those of the USPTO. For this reason, the PTO is seeking greater harmonization and cooperation with these offices. *See The Website of the Trilateral Co-operation*, http://www.trilateral.net/ (last modified Dec. 19, 2007) (noting that the majority of all patent applications worldwide are processed by one of these three offices).

[117] *Phillips*, 415 F.3d at 1317 ("[T]he prosecution history represents an ongoing negotiation between the PTO and the applicant . . ."); Mark A. Lemley & Kimerly A. Moore, *Ending Abuse of Patent Continuations*, 84 B.U. L. REV. 63, 66-67 (2004).

[118] *See* Lemley & Moore, *supra* note __, at 66-68 (describing the prosecution process).

[119] 35 U.S.C. § 102 (2000).

U.S.C. § 102, is actually a list of situations that are deemed to place the invention outside of the patent applicant's ownership, including the fact that another invented the idea before the applicant or that that it was published more than one year before the application was filed.[121] The non-obvious requirement, contained in 35 U.S.C. § 103, is a rule that requires a patentable invention to be more than an obvious advancement on information already in the public domain.[122] Often, obviousness rejections are made by combining several pieces of "prior art" to arrive at the applicant's invention.[123]

The examiner identifies novelty (102) and non-obvious (103) issues in a document called a "rejection," which is exactly what it sounds like: a denial of claims.[124] It is a multipage listing of the rationale for rejection, including the identification of specific pieces of prior art that create a barrier.[125] In response to the rejection, the patent applicant may respond with amendments to the claims or arguments intended to traverse the examiner's objections.[126] Subsequently, a final rejection may issue, which the patentee has yet another chance to traverse or appeal.[127] If a patent eventually issues, one can conclude that the cited art was close enough to the invention to create a problem, but not so close as to preclude patentability. In other words, the patent examination process inherently highlights the most connected patents, which is a subset of all of the patents cited during the prosecution of the application.

With this information in mind, it is relatively simple to construct a system for identifying a relevant subset of patent citations. One can merely look through any rejections in the file history that contain 102 or 103 arguments and list the patents. These patents can form an extremely relevant network of rights.

Importantly, the difference between total citation analysis and 102/103-citation analysis can be considerable. It would manifest itself as fewer network connections and a more accurate picture of the patent environment. For illustrative purposes, consider the following graphic of a set of primary patents with all of their backward citations in contrast to the same patents with only 102/103 citations.[128]

---

[120] 35 U.S.C. § 103 (2000).

[121] 35 U.S.C. § 102(a),(f) (2000).

[122] KSR Intern. co. v. Teleflex, 550 U.S. 398, 127 S.Ct. 1727, 1740-41 (2007) (discussing the rationale for the doctrine).

[123] MPEP § 2141 (8th ed., rev. 7, 2008).

[124] 37 C.F.R. § 1.104 (2007); MPEP § 706 (8th ed., rev. 7, 2008).

[125] 37 C.F.R. § 1.104 (2007) ("In rejecting claims for want of novelty or for obviousness, the examiner must cite the best references at his or her command.").

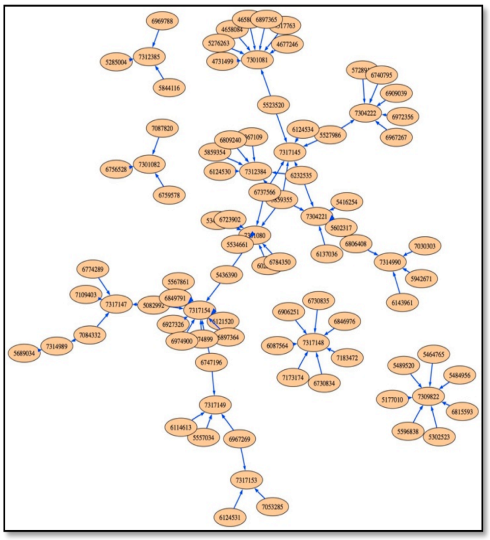
[126] MPEP § 706 (8th ed., rev. 7, 2008).

[127] MPEP § 706.07 (8th ed., rev. 7, 2008).

[128] The citations are a small group of patents used to illustrate the point. True patent networks are extremely detailed and may make use of thousands of patents. *See*, *e.g.*,

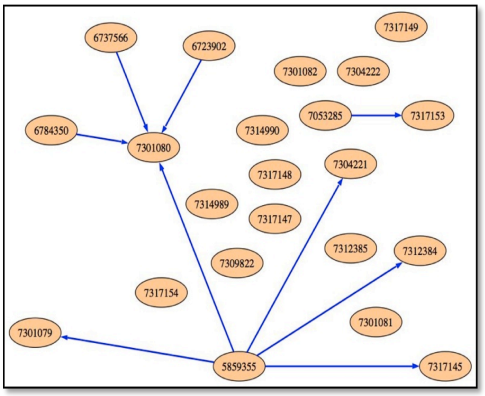
Figure 1

All Backward Citation Links
Patents Issued Nov.-Dec. 2007: Class 800/300.1



Source: USPTO Issued Patents

102/103 Backward Citation Links
Patents Issued Nov.-Dec. 2007: Class 800/300.1



Source: USPTO Issued Patents

A network composed of patent linked by 102/103 rejections should provide better information on which patents may intersect, indicating greater monopoly control or the potential for reducing innovation to some extent.

Another point about 102/103-citation analysis is worth mentioning: it is less open to manipulation than general citation counts. As detailed above, a patent applicant may submit an unlimited number of pieces of prior art in the course of the examination, regardless of whether they have a strong connection to the patented subject matter. Given that some empiricists use citation counts to inform a patent value determination,[129] it is not unreasonable to presume that some patent applications intentionally submit an increased amount of higher art. On the other hand, a patent examiner applies only the 102/103 rejections that are justified by the prior art. The patent applicant has essentially no influence on this process, and it cannot be gamed.

Given the straightforward nature of 102/103-citation analysis and its obvious advantages, one might ask why it has not been employed in the past. The answer lies in the historical difficulty in accessing the information. In order to obtain a patent file history, one formerly had to

---

Clarkson, *supra* note __.

[129] *See supra* note 72, and accompanying text.

travel to the PTO offices in Virginia and request a paper copy, which one could then review and photocopy. Alternatively, one could request a certified copy by mail. Needless to say, for more than a few file histories, this process was inefficient and expensive. However, as a result of the 1999 American Inventors Protection Act,[130] every document in file histories for patents filed after 2000 is available on line.[131] This creates a new opportunity to enhance a well-trod avenue of patent empirics.

B. Empirical Evidence of Increased Relevance of 102/103 Citations

If indeed 102/103-citation analysis identifies relevant connections between patents, empirical evidence should bear this out. Since, as described above, one of the primary qualities that impacts the patent environment is the density of related patents, 102/103-citation networks would be expected to provide a better indication of density. Therefore, this article presents the results of a basic empirical assessment of 102/103-citation density as compared to full citation density in a selection of patents already known to be dense versus one that should be significantly lighter.

The primary hypothesis derived from the above discussion is that there should be a significantly greater number of 102/103 citations per patent in a selection of densely grouped patents as compared to a more dispersed grouping. Notably, research has already demonstrated that the number of overall citations increases in patent groupings that are likely to be dense, such as litigated patents.[132] One would expect 102/103 citations to generally comport with these results. Of course, if 102/103-citation analysis is actually more effective that general citation analysis, one would expect to see significant differences even when general citation is indistinct. This is a secondary hypothesis that was investigated.

In order to test the effectiveness of the 102/103-citation analysis, a grouping of patents that is demonstrably dense first must be identified. As noted above, it is extremely time consuming (arguably impossible) to conclusively prove density through qualitative analysis of patents.[133] However, the literature describes at least two possible proxies that exist for density. One is the existence of patent "pools," which are co-licensing arrangements made by competitors to overcome a thicket problem.[134] Another is the increased likelihood that a patent will be litigated.[135] This

---

[130] Pub. L. No. 106-113, § 4712, 113 Stat. 1501 (codified as amended in scattered sections of Title 35 of the U.S. Code).

[131] *See* USPTO, *PAIR FAQ*, http://www.uspto.gov/ebc/pair/pair_faq.html.

[132] *See*, *e.g.*, Allison et al, *supra* note __, at 454.

[133] *See supra* notes 39-44, and accompanying text.

[134] Shapiro, *supra* note __, at 134.

[135] *See* Allison et al, *supra* note __, at 439-34 (explaining the connection between

article assesses 102/103-citation analysis using both proxies.

One limitation of the studies below that should be noted at the outset is that it is skewed toward recent patent. This is due to the fact that USPTO on-line file histories exist only for patents *filed* after 2001.[136] Such patents generally issued, at the earliest, in 2004-2005.[137] However, there is no reason to believe that reliance on recently issued patents has any impact on the analytical framework. The primary difference one expects with recent patents is greater overall citation[138] and the fact that forward citation will be less complete.

[AUTHOR NOTE: THE PURPOSE OF THIS DRAFT IS TO DISCUSS THE IDEA OF ENHANCED CITATION ANALYSIS. ADDITIONAL DATA COLLECTION AND ANALYSIS ARE CLEARLY NECESSARY FOR EMPIRICAL PROOF. THIS WORK IS ONGOING]

1. Validation with Density Based on Increased Incidence of Patent Pooling

A dense group of patents was identified using technology classifications employed by the PTO's Manual of Classification (MOC).[139] Although it is generally understood that PTO classifications are somewhat arbitrary and may be both under inclusive and over inclusive at the same time with respect to a technology field, the system provides a useful first step toward categorization.

The dense patents were derived from a patent classification that has been subject to pooling. It has been argued in the literature that patent pools are likely coextensive with patent thickets.[140] One of the most famous patent pools that has been described in the literature is the MPEG pool.[141] This collection of patents is deemed necessary to produce devices and

---

litigation and patent value, and the rationale for the assumption that characteristics of litigated patents are different than non-litigated patents).

[136] In the case of highly litigated and visible patents, older file histories may also be collected and available through services like Westlaw's file history database (us-pat-history).

[137] Patent pendencies tend to be about three or more years. *See* Andrew W. Torrance, *Patenting Human Evolution*, 56 KAN. L. REV. 1075, 1092 n.66 (2008)

[138] John R. Allison & Mark A. Lemley, *The Growing Complexity of the U.S. Patent System*, 82 B.U. L. REV. 77, 101-103 (2002).

[139] The Manual of Classifications is available in electronic form. http://www.uspto.gov/go/classification/.

[140] *See*, *e.g.*, Clarkson, *supra* note __, at 14-15; Shapiro, *supra* note __, at 134.

[141] *See*, *e.g.*, Clarkson, *supra* note __, at 20-21; Joseph Scott Miller, *Standard Setting, Patents, And Access Lock-In: RAND Licensing and the Theory of the Firm*, 40 IND. L. REV. 351 (2007) (discussing the MPEG pool and its interaction with the U.S. government).

software that utilize the MPEG architecture.[142] Ideally, one would simply use the patents involved in the pool itself for use in the study. Unfortunately, there are an insufficient number of recent patents in the MPEG pool to provide a study set. As a reasonable substitute, thirty patents were collected from one of the classes involved in the MPEG pool, under the assumption that this class is likely to be denser than others. Described as high-density control concerned bandwidth attenuation in digital communications, it is numbered class 375/250 (Dig).

For a comparison group, thirty patents with a low likelihood of overlap were identified. Maintaining the same idea of using PTO patent classifications, a class was used that contains almost entirely discrete inventions in the form of individual plant varieties. Described as genetically modified corn varieties, class 800/300.1 (Corn) is home to many patents on individual germplasm that do not need to interact with other patents in the class.

Following the identification of the respective patent numbers above, citations for each patent were collected using the PTO's online database.[143] In order to determine which of these citations formed the basis of a 102 or 103 rejection by a patent examiner, electronic file histories — the record of each document filed in the prosecution of the patent — for each patent were accessed through the PTO's public Patent Application Information Retrieval (PAIR) system.[144] Within each file history, the examiner's rejections (if any) were reviewed and any 102/103 cites recorded.

The results of the patent pooling study suggest that 102/103-citations signal the density of the digital bandwidth attenuation class, but general patent citations do not (*see* Table 2, below).

---

[142] Introduction to MPEG-LA's AVC Patent Portfolio License, http://www.mpegla.com/avc (last visited Jan. 9, 2009).

[143] USPTO Patent Full-Text and Image Database, http://patft.uspto.gov/netahtml/PTO/search-bool.html (last visited Jan. 9, 2009).

[144] USPTO, Patent Application Information Retrieval, http://portal.uspto.gov/external/portal/pair.

Table 2

102/103 Citations Per Patents

	Mean 1	Mean 2	Var. 1	Var. 2	Prob 2-Tail
Corn v. Dig	.74	3.23	1.59	7.71	.00005
Corn v. Dig (Corrected)[145]	.74	3.26	1.59	7.93	.00003

All Citations Per Patents

	Mean 1	Mean 2	Var. 1	Var. 2	Prob 2-Tail
Corn v. Dig	8	13.87	106.2	405.85	.155
Corn v. Dig (Corrected)	8	10.73	106.2	104.13	.302

There were approximately 0.74 102/103 citations per patent in the genetically modified corn group as compared to 3.23 citations per patent in the pooling class. Statistical analysis using a 2-tailed t-test assuming unequal variances demonstrates that these differences are highly significant. In contrast, there were approximately 8.0 general citations per patent in the genetically modified corn class versus 13.87 citations per patent in the digital bandwidth attenuation class. The same statistical analysis demonstrates that this difference is not significant.

The results appear to confirm both hypotheses. However, it should be noted that there is some potential for skewing of the results in these samples; factors such as common ownership of patents and other characteristics may have the effect of exaggerating unusual prosecution or examination behavior, and this could have affected the results. Thus, in order to further validate 102/103-citation analysis, another measure is helpful.

2. Validation with Density based on Involvement in Litigation

A useful second proxy for density is a patent's involvement in litigation. It is intuitive that a patent is more likely to exist in a dense environment if it is involved in litigation.[146] This is due to the fact that litigated patents tend to exist in areas of high economic interest, which in

[145] One patent had a such a large number of citations that the study was run again without it to ensure that the results were not unduly skewed.

[146] Note that most patents are not involved in litigation, so this is truly a narrow subset.

turn attracts investment in innovation leading to more patenting. In accord with this intuition, Allison et al., demonstrated that litigated patents have a higher incidence of overall citation.[147] A measure of high density is reasonably supplied by a selection of litigated patents. Conversely, a measure of lower density may be supplied by selecting non-litigated patents from the general population.

Recent litigated patents were identified using the Stanford IP Litigation Clearinghouse.[148] This database attempts to comprehensively record patent litigations across the country. For the present study, suits filed in 2008 were surveyed, and patents were collected that met the time limitation described above (generally those with numbers above Pat. no. 6,800,000). The total number of patents collected was 78. To minimize regional effects, approximately 1/3 of the patents were collected from litigations in the U.S. District Court for the Central District of California, 1/3 from the U.S. District Court for the Southern District of New York, and 1/3 from the U.S. District Court for the Northern District of Illinois.

Non-litigated patents were collected from two groups. The first was a random selection of patents[149] from a classification that is most often involved in litigation according to the Stanford IP Litigation Clearinghouse: class 514, "Drug, bio-affecting and body treating compositions," (highly litigated).[150] One would hypothesize that this selection of patents might be slightly denser than a totally random sample of patents. One hundred patents were collected from this group. For the sake of comparison, a group of ninety-four patents was collected from an emerging technology field, second-generation biofuels (Bio) that the author demonstrated to be dispersed (i.e., not dense) in a prior paper.[151] In both groups as compared to the actually litigated patents, one would expect to see a difference if patent density is being measured at a meaningful level.

As in the validation study related to patent pooling, following the identification of the respective patent numbers above, citations for each patent were collected using the PTO's on-line database. To determine which of these citations formed the basis of a 102 or 103 rejection by a patent examiner, electronic file histories for each patent were accessed through public PAIR system. Within each file history, the examiner's rejections (if any) were reviewed and any 102/103 cites recorded.

The results of the litigation study suggest that 102/103-citations signal the higher density of the actually litigated patents (*see* Table 3, below).

---

[147] *See*, *e.g.*, Allison et al, *supra* note __, at 454.

[148] http://lexmachina.stanford.edu (last visited Jan. 8, 2009).

[149] Starting from the last patent issued in 2008 and proceeding in reverse order, every 150 patents in the class was selected until 100 were obtained.

[150] http://lexmachina.stanford.edu/patents/byclass

[151] *See* Cahoy & Glenna, *supra* note __.

Table 3

102/103 Citations Per Patents

	Mean 1	Mean 2	Var. 1	Var. 2	Prob 2-Tail
High-Lit vs. Act Lit	.75	1.92	1.83	3.29	.00000
Bio vs. Act. Lit	.73	1.92	1.51	3.29	.00000
Log-Trans. High-Lit vs. Act Lit	.08	.27	.04	.08	.00000
Log-Trans. Bio vs. Act. Lit	.09	.27	.03	.08	.00000

There were approximately 0.74 102/103 citations per patent in the group from the highly litigated class and 0.73 citations per patent in the Bio group, as compared to 1.92 citations per patent in the group of patents that were actually litigated. Statistical analysis using a 2-tailed t-test assuming unequal variances demonstrates that these differences are highly significant. To ensure that outlier patents did not unduly skew the data,[152] it was log transformed and the subjected to statistical analysis. The significance remained.

In contrast to the patent pool-related study, the results of the litigation study suggest that general citations also signal the higher density of the actually litigated patents (*see* Table 4, below).

[152] For example, one patent had over 500 prior art citations, while most other had 60-80.

Table 4

All Citations Per Patents

	Mean 1	Mean 2	Var. 1	Var. 2	Prob 2-Tail
High-Lit vs. Act Lit	10.90	70.73	192.67	18396.35	.00022
Bio vs. Act. Lit	8.84	70.73	177.15	18396.36	.00014
Log-Trans. High-Lit vs. Act. Lit	.73	1.35	.31	.46	.00000
Log-Trans. Bio vs. Act. Lit	.64	1.35	.27	.46	.00000

There were approximately 10.9 general citations per patent in the group from the highly litigated class and 8.8 citations per patent in the Bio group, versus 70.73 citations per patent in the in the group of patents that were actually litigated. Statistical analysis using a 2-tailed t-test assuming unequal variances demonstrates that these differences are highly significant. As above, to ensure that outlier patents did not unduly skew the data, it was log transformed and the subjected to statistical analysis. Again, the significance remained.

3. Discussion of Study Results

The empirical data generally support the hypothesis that 102/103-citation analysis is useful for providing information regarding the relationship between patents. Higher citation numbers appear to coincide with environments in which one would expect to find a higher density of patents. This is consistent with the notion that an examiner will be inclined to issue more rejections on the basis of lack of novelty or non-obviousness in a very crowded art. The analysis appears to be an important snapshot of the patent landscape.

However, the hypothesis that 102/103-citation analysis is actually better than general citation analysis in predicting density cannot be entirely confirmed. In the litigation-related patents, general citation is arguably a better indicator. The data at least suggest that 102/103-citation analysis provides information about patent landscapes that general citation obscures, serving as an important complement.

In addition, there are indications that other patent characteristics or sample size may impact the results. For example, although the group of

actually litigated patents demonstrated a higher 102/103-citation count per patent than the highly litigated or biofuel patents, there were differences within the sample between patents from the respective district courts.[153] This suggests that the type of patents involved in litigation may make a difference (*i.e.*, there could be significant density differences within litigated patents as a group). Also, not presented above are the results of a survey of a smaller sample of a class of patents with lower litigation rates (class 242, "winding, tensioning or guiding"). This survey demonstrated 102/103-citation rates that were lower, but not significantly so, than the group of actually litigated patents. While this may simply mean that the class is dense regardless of litigation incidence, another confounding factor may be involved. Again, these issues suggest the use of 102/103-citation analysis as an important complement to rather than a replacement of traditional citation analysis.

CONCLUSION

Patent citation analysis has proven itself a very valuable tool. The use of only those citations that served as the basis of a formal patentability rejection during the examination provides a very promising possibility for increasing citation relevance. Such citation links are relevant by definition, and they are directed to the core of the patented subject matter. New electronic means of accessing patent examination histories has opened the door to this analysis, and it is finally possible to assess its viability as an analytical tool. Although 102/103-citation analysis may not yet be as easy to employ as standard citation methods, theoretical assessment and empirical evidence suggest that it is worth the effort. When the impact on business is so important, searching away from the metaphorical streetlamp is necessary.

* * *

[153] Patents collected from litigations in the U.S. District Court for the Southern District of New York had a relatively low 102/103-citation count, while patents from litigations in the Northern District of Illinois had a very high count.