ALL THE WORLD'S A NETWORK

A COMPREHENSIVE GLOBAL PATENT CITATION TREE

Andrew W. Torrance[1] & Jevin D. West[2]

I. INTRODUCTION[3]

Patent rights are granted by almost all governments around the world. A primary motivation for these grants is to foster technological innovation. By offering inventors the prospect of limited rights to exclude others from exploiting their inventions, patents are often assumed to spur inventive activity, followed by commercialization of new inventions. This view of patents as important policy levels for spurring innovation has existed for a long time. In fact, one of the first successful global treaties was the Paris Convention for the Protection of Industrial Property ("Paris Convention"), which offers mutual recognition of patents, designs, and trademarks, came into force on July 7, 1884. One of the foundational principles of the Paris Convention is the international interconnectedness of patent documents.[4]

Over centuries, an enormous amount of data associated with patenting and patent systems has accumulated in patent offices, authorities, and institutions around the world. Among the categories of data carefully maintained by patent offices is prior art. Prior art consists of documents or artefacts that precede the priority date of a patent or patent

[1] Earl B. Shurtz Research Professor, University of Kansas School of Law; Visiting Scholar, Sloan School of Management, Massachusetts Institute of Technology; Research Fellow, Gruter Institute for Law & Behavioral Research. Prof. Torrance wishes to thank Nathan Mannebach, Max McGraw, Edgar Acevedo-Pando, J.B. Fitzgerald, Aaron Vanderpool, and Daniel Kopp for their excellent research assistance.
[2] Assistant Professor, DataLab, Information School, University of Washington.
[3] Part of this section has been adapted from Torrance, A.W., and West, J.D.. (2017) All Patents Great and Small - A Big Data Network Approach to Patent Valuation, Virginia Journal of Law and Technology, Vol. 20, No. 3, pp. 466-504.
[4] We use "patent document" to refer to any of a variety of documents representing patent rights, including a patent, patent application, Patent Cooperation Treaty ("PCT") application, and European Patent prior to nationalization.

application, and are relevant to the patentability of its claims. Patent applications, and the patents they may become,[5] cite prior art both during prosecution and on their faces once granted. These patent citations link patent documents to each other in a time-directed manner. Moreover, what begins as a newly-filed patent application that cites prior art often later itself becomes prior art for subsequent patent applications. These patent citations form networks, including a worldwide patent citation network that encompasses patent documents from almost every country.

Measuring patent importance is a major goal of scholars in both patent law and patent economics. However, doing so objectively, accurately and consistently has proved exceedingly difficult. At least part of the reason for this difficulty is that patents themselves are complex documents that are difficult even for patent experts to interpret. In addition, issued patents are the result of an often long and complicated negotiation between applicant and patent office (*e.g.*, United States Patent & Trademark Office ("USPTO"), European Patent Office ("EPO"), Canadian Intellectual Property Office ("CIPO")), the result of which is an opaque "prosecution history" upon which depend the scope of claimed patent rights. In this Article, we use the relative positions of patents and patent applications embedded within a comprehensive patent citation network to measure the importance of those patents within the network. We tend to refer to the "importance" of patents instead of "value", but there is good reason to believe that these two concepts share a very similar meaning.[6]

Patents are not merely isolated descriptions of inventions deemed new and useful enough to warrant government *imprimatur*. On the contrary, patents frequently cite other patents and references (*e.g.*, scientific articles, webpages, datasets)[7] and therefore are more

---

[5] Hereafter, we use "patent" to denote a granted or issued patent or a patent application. When we intend to refer specifically to either a patent or patent application, we indicate this.
[6] See, generally, Torrance, A.W., and West, J.D.. (2017) All Patents Great and Small - A Big Data Network Approach to Patent Valuation, Virginia Journal of Law and Technology, Vol. 20, No. 3, pp. 466-504.
[7] A patent may also make reference to a physical artifact, such as a commercial product, or processes capable of being carried out in the physical world. For example, in U.S. patent law, under 35 U.S.C. §§102 and 103, physical objects or processes can constitute prior art capable of potentially anticipating or rendering obvious a patent claim.

than mere collections of isolated documents. World-wide, tens of millions of patents are interconnected by hundreds of millions of citations. Patents and the citations that interconnect them form a vast network, with patents as "nodes" and citations as "links" among them. This "patent citation network" represents the aggregation of hundreds of millions of deliberate choices individual patent applicants, or patent attorneys or agents representing them, and patent examiners have made about how to situate their inventions in relation to others' inventive ideas.

The structure of this network contains a wealth of information about the patents, and the communities within which the patents reside. We use eigenvector centrality and hierarchical clustering methods to evaluate the patent citation network of all patents worldwide found in the spring 2017 PATSTAT database.[8] As noted above, and explained in detail below, patent importance is measured as a property of a patent's position within the patent citation ecosystem.

II. PATENT IMPORTANCE[9]

A. PATENTS AND VALUE

An effective method of patent valuation has consistently eluded patent scholars and practitioners, but not for a lack of effort. In fact, determining an accurate method of estimating patent value is something of a Holy Grail within patent studies and practice. Approaches ranging in complexity from the "rule of thumb", which arbitrarily divides licensing profits in a 25/75 split, to the Black-Scholes equation, which is more commonly used to value stock market options, have been applied to the problem, but none have

---

[8] We rely primarily on the bibliographic data found in the EPO's spring 2017 PATSTAT database, released in March 2017 (https://www.epo.org/searching-for-patents/business/patstat.html#tab-1). There may patents from this time period that are absent from our patent citation network. If so, the explanation is that the EPO has not made sufficient data about these patents available through the spring 2017 PATSTAT database. The gigantic size of this collection of references makes verifying the perfect completeness of our patent data set mathematically difficult. Despite this *caveat*, we believe our collection of patents is among the most complete available.

[9] Adapted from Torrance, A.W., and West, J.D.. (2017) All Patents Great and Small - A Big Data Network Approach to Patent Valuation, Virginia Journal of Law and Technology, Vol. 20, No. 3, pp. 466-504.

satisfied the patent economics community.[10] All of the proposed approaches generally fit into two categories: financial valuation methods and non-financial valuation methods. The method used in this Article is non-financial. Nevertheless, we provide a brief overview of other approaches to patent valuation to place our method in context.

B. FINANCIAL PATENT VALUATION METHODS

The literature on patent valuation consistently divides financial methods into three main categories of increasing complexity: cost, market, and income methods.

1. COST METHODS

The cost method values a patent asset by calculating the cost of replacing it, reconstructing it, or substituting it for another asset, and then equating that cost to the value of the new asset.[11] Simply knowing how much the licensor spent creating the patent is not enough, however, because the licensee could be a more efficient innovator, and the patent landscape would have changed from the time of invention to the time of valuation.[12] The cost method does not take into account other competitors in the market, any future benefits possibly derived from taking advantage of the patent, or the economic life of the patent, and those are but a few of the disadvantages of this method.[13] Input costs alone tend not to be good indicators of patent value because many individual inventors accidentally invent products or processes protected by extremely valuable patents, while many large, wealthy firms fail to develop valuable patents despite the investment of prodigious sums on research and development.

---

[10] GOLDSCHEIDER, LICENSING AND THE ART OF TECHNOLOGY MANAGEMENT 11–17, sec. 11.4 (Thomson Reuters, Westlaw, 2011); F. Russell Denton & Paul J. Heald, Random Walks, Non-Cooperative Games, and the Complex Mathematics of Patent Pricing, 55 Rutgers L. Rev. 1175, 1177 (2003) ("[T]he Denton Variation of the Black-Scholes equation, exploits the similarities between the option to buy stock and the option to develop an invention.").
[11] Technology Licensing and Development Agreements § 6.8.1 (Matthew Bender, Rev. Ed. 2015) [hereinafter *Bender Treatise*].
[12] *Id.*
[13] *Id.* at § 6.8.2.

2. MARKET METHODS

The standard market method is another relatively straightforward valuation technique that involves using historical prices agreed upon for the subject patent asset, and then making adjustments based upon the current patent landscape, as well as the particular market needs of a new license.[14] Another, indirect, version of this method consists of finding similar technologies that have already been valued, and then basing estimation of a patent on the values of these similar technologies.[15] Both parties to a patent licensing negotiation are usually familiar with the subject patent's technological field, and, consequently, tend to be comfortable with this valuation method; however, unlike the housing market, there can often be substantial differences among even similarly-situated patent assets, which can confound the comparability of putatively similar patent assets.[16]

Another useful variation of this method, which is made possible when accurate historical information exists as to patent pricing, is the rating and ranking method because it quantifies the value differences between the subject patent and well-characterized patents.[17] The quantification of this difference is done by using a set of factors, generally the Georgia-Pacific factors[18], and analyzing how the subject patent compares in value to that of patents with known values.[19] If, after analyzing all 15 enumerated Georgia-Pacific factors, the subject patent outperforms the patents of known value, then the subject patent will tend to be valued more highly than the patents of known value, and *vice versa*.[20] One commentator even takes the unique approach of combining the rating and ranking method with non-financial indicators, such as payment of maintenance fees and technology class,

---

[14] Michele Floyd & Lawrence Wu, THE REVOLUTION IN THE LAW AND ECONOMICS OF ANTITRUST CLASS CERTIFICATION § 3.03 (LexisNexis, Inc., 2015).
[15] *Id.*
[16] *Id.*
[17] Richard Razgaitis, Pricing the Intellectual Property of Early-Stage Technologies: A Primer of Basic Valuation Tools and Considerations 830 (2007).
[18] Georgia-Pacific Corp. v. United States Plywood Corp., 318 F. Supp. 1116, 1119–20 (S.D.N.Y. 1970), modified and aff'd, 446 F.2d 295 (2d Cir.).
[19] Razgaitis, supra note 8 at 831.
[20] *Id.*

to rank patents against each other to assign them a comparative value.[21] The problems of identifying patent assets of known value for comparison, deciding which comparative factors to use, and knowing how to rank the patent assets in light of each of those factors can be very challenging to solve, but this method at least provides approaches for quantifying patent assets.

3. INCOME METHODS

The final financial method for patent valuation is the income method, which is widely considered the most complex, but also the most economically-suitable, approach.[22] This method is based on the "assumption that the value of any asset can be expressed as the present value of the future stream of economic benefits that can be derived from its ownership."[23] To carry out this method, an interested party projects the cash flow a patent asset will earn for that party over the expected lifetime of that asset, that final value is offset by a discount rate that accounts for the interest rate and degree of risk, and finally that patent asset value is reduced to a present value.[24] This is yields a discounted cash flow.[25] There are many variables in this calculation, any of which may introduce calculation errors, though various income methods have been developed to account for those variables, including discounted cash flow, real options, binomial expansion, and Monte Carlo methods.

Discounted cash flow is the simplest method, but has two major, though subtle, variations. The first is a method that uses patent claim analysis to achieve a more accurate projected revenue stream.[26] The originators of this variation advocate a method in which deciphering the patent claims informs a company about which products are covered by

---

[21] Jonathan A. Barney, A Study of Patent Mortality Rates: Using Statistical Survival Analysis to Rate and Value Patent Assets, 30 AIPLA Q. J. 317 (2002).
[22] *Bender Treatise*, supra note 2 at § 3.03.
[23] F. Russell Denton & Paul J. Heald, Random Walks, Non-Cooperative Games, and the Complex Mathematics of Patent Pricing, 55 Rutgers L. Rev. 1175, 1188 (2003).
[24] *Id.*
[25] Razgaitis, *supra* note 8 at 839.
[26] Malcolm T. Meeks & Charles A. Eldering, Patent Valuation: Aren't We Forgetting Something? Making the Case for Claims Analysis in Patent Valuation by Proposing a Patent Valuation Method and a Patent-Specific Discount Rate Using the CAPM, 9 Nw. J. Tech. & Intell. Prop. 194, 234 (2010).

those claims.[27] Knowing which products fall within a patent's claims allows a company more accurately to project the revenue associated with that patent.[28] Finally, that revenue stream is discounted at a patent-specific discount rate using the capital asset pricing model.[29] The other variation attempts to find future cash flows from a cost-reducing technology covered by a patent's claims by adding together the cash flows gained from competitive advantage, licensing income, and maintenance costs of the patent.[30] This variation on discounted cash flow only works with patents whose claimed technologies have already been well developed, and, thus, do not require substantial additional investments.[31]

When substantial investments are required, real option valuation based on the Black-Scholes equation is more apt.[32] The Black-Scholes equation was created to predict company revenues in order to properly value stocks.[33] Similarly, to properly value a patent, company revenues gained from that patent must be accurately predicted.[34] Denton and Heald suggest modifying the Black-Scholes equation to take advantage of "similarities between the option to buy stock and the option to develop an invention" such as "definite expiration dates and sequentiality of investment moments" to make patent valuations.[35] The major advantage of a real option valuation is that it allows for the possibility that a company will abandon an invention once it becomes clear that invention will not be profitable, allowing mitigation of risk.[36]

---

[27] *Id.*
[28] *Id.*
[29] *Id.* at 224–25.
[30] Sander van Triest & Wim Vis, Valuing patents on cost-reducing technology: A case study, 105 Int. J. Production Economics 282, 283 (2007).
[31] *Id.* at 284.
[32] *Id.*
[33] Denton & Heald, *supra* note 14 at 1176.
[34] *Id.*
[35] *Id.* at 1176–77.
[36] *Bender Treatise*, *supra* note 2 at § 6.4 ("Real options treats risk differently than income method. The latter uses a discount premium rate to reduce expected income, whereas real options considers that the manager can dramatically reduce risk by making choices and using judgment as time goes by.").

Binomial expansion is "a more advanced application of [r]eal [o]ption [v]aluation where there exists 'options on options'."[37] This allows a company to differentiate possible outcomes by "milestone" events, because at each one of these events the company can assign the likelihood of each outcome.[38] Although a single forecast takes these milestone events into consideration, breaking them out into a decision tree allows for more transparency as well as further analysis of the most critical valuation issues.[39]

Where binomial expansion only allows for binary outcomes of set probabilities, the Monte Carlo technique takes advantage of this result by simulating thousands of scenarios over different probability ranges.[40] For example, when the input costs for a given scenario are equally probable between $1 million and $3 million, the binomial expansion method would have to choose two numbers within that range, but the Monte Carlo technique allows the likelihood of every possibility in that range to be calculated.[41] The outcome of the simulation is a confidence interval of the most likely values, which gives the estimated worth of the patent being analyzed.[42] Some researchers have extended this method even further by using a sensitivity model to demonstrate how a value varies with the model's parameters because of the difficulty in adjusting for the appropriate discount rate.[43]

C. NON-FINANCIAL PATENT VALUATION METHODS

1. FORWARD AND BACKWARD CITATION METHODS

A large number of established, non-financial indicators of patent value exist including forward citations, backward citations, family size, number of claims, key inventors, and market value of corporation among others.[44] Forward and backward

---

[37] BUSINESS SPREADSHEETS, http://www.business-spreadsheets.com/help.asp?t=21.
[38] Floyd & Wu, *supra* note 5.
[39] *Id.*
[40] *Id.*
[41] Razgaitis, supra note 8 at 852.
[42] Gary S. Stacey, Valuing Intellectual Property, Technology Strategies 22–25 (1989).
[43] Jow-Ran Chang, Mao-Wei Hung & Feng-Tse Tsai, Valuation of Intellectual Property: A real option approach, 6 Journal of Intellectual Capital 339, 353 (2005).
[44] Markus Reitzig, Improving Patent Valuations for Management Purposes—Validating New Indicators by Analyzing Application Rationales, 33 Research Policy 939, 941 (2004).

citations are the most studied, and, generally, the best validated of those these.[45] Considerable research suggests that the numbers of forward and backward citations associated with a patent are positively correlated with the value of that patent.[46] One recent study, relying on a confidential corporate dataset, has questioned how reliable citations are as indicators of value above a threshold of citations.[47] However, the weight of evidence spanning the past three decades robustly suggests that patent citations can be powerful indicators of patent value. Furthermore, as explained below, the method of weighting individual patent citations used in our analysis is especially comprehensive and accurate.

2. PATENT CITATION NETWORK METHODS

[45] *See, e.g.*, Francis Narin, Kimberly S. Hamilton, & Dominic Olivastro, The Increasing Linkage between U.S. Technology and Public Science, 26 Research Policy 317, 317 (1997); Bronwyn H. Hall, Adam B. Jaffe, & Manuel Trajtenberg, Market Value and Patent Citations: A First Look, National Bureau of Economic Research (2000) (discussing different aspects of using patent citations).

[46] *See* Dietmar Harhoff, Frederic M. Scherer, & Katrin Vopel, Citations, Family Size, Opposition and the Value of Patent Rights, 32 Research Policy 1343, 1359–60 (2003); Dietmar Harhoff, Francis Narin, F.M. Scherer & Katrin Vopel, *Citation Frequency and the Value of Patented Inventions*, 81 Rev. of Econ. and Stat. 511 (1999) (This study involves a survey of 964 inventions made in the U.S. and Germany, and on which German patent renewal fees were paid to full-term expiration in 1995 estimated economic value of the patents. After considering patent citations, patents renewed to full-term were significantly more highly cited than patents allowed to expire before their full term. The higher an invention's economic value estimate was, the more the patent was subsequently cited. Patents were reported to be relatively valuable by the companies holding them are more heavily cited in subsequent patents. A two-stage relationship between economic values and citation counts was observed: first, patents that are renewed to full-term expiration in environments such as Germany, with highly progressive annual maintenance fees, are more highly cited than run-of-the mill patents allowed to expire before running to full term, and full-term patents are more valuable on average than patents allowed to lapse at midterm: second, within the relatively select cohort of full-term patents, citation frequency rises noisily with reported economic value. The method employed involved hypothesizing that more valuable patents are more frequently cited, focusing on the private value of their survey patents and the patents' underlying inventions to patent holders, rather than their social value. Germany was chosen because it is one of the most progressive patent renewal fee systems in the world. Telephonic contacts were achieved in Germany with the holders of 1,352 patents. The authors tried to link the German patents to related U.S. patents, but attrition occurred because not all German patents disclosed prior U.S. applications. There were 485 patents that were parallel and linked. The authors employed an asset-value approach.)

[47] David S. Abrams, Ufuk Akcigit, & Jillian Popadak, Understanding the Link between Patent Value and Citations: Creative Destruction or Defensive Disruption?, University of Pennsylvania and NBER (April 8, 2013).

The method of analysis proposed in this paper is an improvement on the patent citation networks already suggested by several academics. Previously, patent citation networks have been shown to approximate "scale-free networks", which are characterized by a few, select hubs through which a large amount of information flows.[48] This network was made using the relatively simple method of counting the number of citations received by each patent and then mapping that information.[49] Further research has revealed that, not only do patent citation networks highlight the most cited patent in each technology field, but also the technological trajectory of the field.[50] Frequently, these citation networks are only used to show trends in a certain technology fields or productivity of certain patents without evaluating their individual importance.[51] This Article improves these techniques by calculating accurate individual patent importances from the patent citation network.

3. LITIGATED PATENT METHODS

Litigated patents tend to possess disproportionately high private value.[52] Building off that assumption, researchers have identified trends in the characteristics litigated patents which can be applied to future patents to determine their value.[53] The researchers confirmed that litigated patents tend to have a greater number of forward and backward citations, but they also found that more claims, longer prosecution time, and larger patent family size were also positively correlated with value.[54] This study suggest that the more time and money a firm invests into patent prosecution, the more likely it is that the

---

[48] Chaomei Chen & Diana Hicks, Tracing Knowledge Diffusion, 59 Scientometrics 199, 201 (2004).
[49] *Id.* at 203.
[50] Adam B. Jaffe & Gaetan de Rassenfosse, Patent Citation Data in Social Science Research: Overview and Best Practices, National Bureau of Economic Research 21–22 (2014).
[51] *See* Id. at 22–23; *see also* Bernard Gress, Properties of the USPTO Patent Citation Network: 1963-2002, World Patent Information (2009).
[52] *See* Dietmar Harhoff, Frederic M. Scherer, & Katrin Vopel, Citations, Family Size, Opposition and the Value of Patent Rights, 32 Research Policy 1343, 1359–60 (2003); John R. Allison et al., Valuable Patents, 92 Geo. L.J. 435, 438 (2004); John R. Allison, Mark A. Lemley, & Joshua Walker, Extreme Value or Trolls on Top? The Characteristics of the Most-Litigated Patents, 158 U. Pa. L. Rev. 1, 5 (2009).
[53] John R. Allison et al., Valuable Patents, 92 Geo. L.J. 435, 438 (2004).
[54] *Id.* at 451–460.

resulting patent will be litigated.[55] The fact that litigated patents have characteristics already proven to be associated with high value lends credence to the assumption that litigated patents themselves are more valuable.

Researchers empirically tested the hypothesis that litigated patents are more valuable by comparing patents that have been litigated once with those that have been litigated eight or more times.[56] If a litigated patents tends to be more valuable than non-litigated patents, a semi-overlapping group of researchers wondered, perhaps the more times a patent is litigated, the more valuable it is. This latter group empirically demonstrated that patents litigated eight or more times tend to possess an even more striking constellation of *indicia* characteristic of valuable patents compared to patents litigated fewer times, especially a single time.[57] Combining this finding with previous research, Allison *et al.* concluded that "the intuitive relationship between value and litigation is indeed the right one."[58]

However, it should be noted that not all studies of litigated patents share the previously mentioned enthusiasm for forward and backward citations as a metric for valuing patents.[59] The studies that made these findings looked not only at what patents were litigated, but also at the outcomes of that litigation, and relied on the reasonable observation that a patent has no value if it is involved in litigation in which a court finds its claims invalid.[60] One study, which compared patents found invalid by a court with those not found invalid in a final adjudication, found that the number of citations a specific patent possesses is negatively correlated with a finding of validity.[61]

---

[55] *Id.* at 461.
[56] John R. Allison, Mark A. Lemley, & Joshua Walker, Extreme Value or Trolls on Top? The Characteristics of the Most-Litigated Patents, 158 U. Pa. L. Rev. 1, 5 (2009).
[57] *Id.* at 28.
[58] *Id.*
[59] *See* Michael Risch, A Generation of Patent Litigation: Outcomes and Patent Quality, 52 San Diego L. Rev. 67, 70. *See also* John R. Allison, Mark A. Lemley & Joshua Walker, Patent Quality and Settlement Among Repeat Patent Litigants,99 Geo. L.J. 677, 681, 686-87 (2011).
[60] Michael Risch, A Generation of Patent Litigation: Outcomes and Patent Quality, 52 San Diego L. Rev. 67, 68 (2015) ("Perhaps the simplest measure of quality is whether a patent is valid, that is, whether it is novel, nonobvious, and otherwise compliant with the Patent Act.").
[61] *Id.* at 118.

In our previous article, *All Patents Great and Small: A Big Data Network Approach to Valuation*, we examined the relationship between whether a U.S. patent has been litigated and its importance as calculated using the ALEF method. Figure 1 depicts our main result.[62]

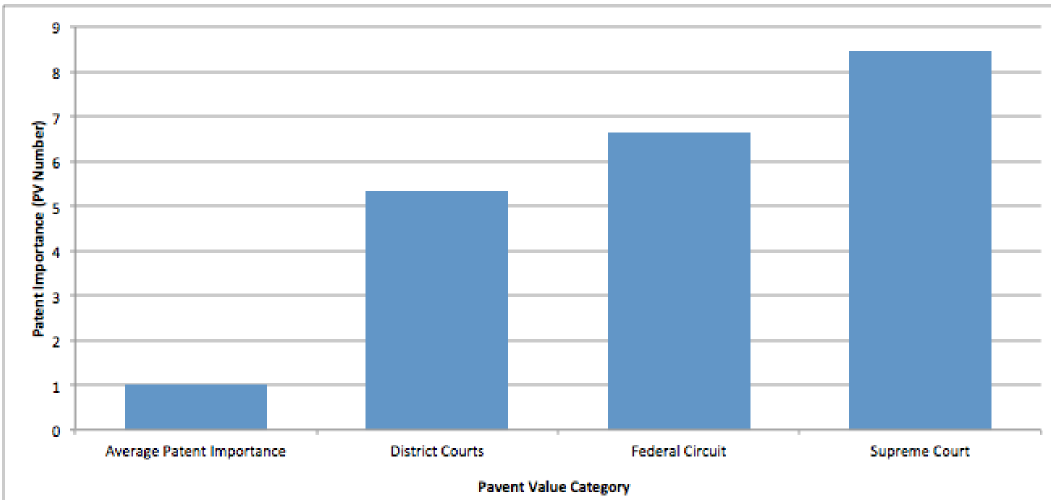


Figure 1: Average patent value for district courts, federal circuit, and supreme court compared to overall average patent importance (1.0).

We found that patent importance tends to be substantially higher for patents litigated in U.S. federal court.[63] Moreover, patent importance rises with level of court: it is lowest when litigated in federal district court (the initial trial court level in the U.S. federal court system); it is higher when litigated in the Court of Appeals for the Federal Circuit ("CAFC")(the first level of federal appeals court for patent cases); and, patent importance is highest when litigated in the U.S. Supreme Court.[64]

III. CITATION NETWORK APPROACHES TO PATENT ANALYSIS

[62] Torrance, A.W., and West, J.D.. (2017) All Patents Great and Small - A Big Data Network Approach to Patent Valuation, Virginia Journal of Law and Technology, Vol. 20, No. 3, pp. 466-504.
[63] Torrance, A.W., and West, J.D.. (2017) All Patents Great and Small - A Big Data Network Approach to Patent Valuation, Virginia Journal of Law and Technology, Vol. 20, No. 3, pp. 466-504.
[64] Torrance, A.W., and West, J.D.. (2017) All Patents Great and Small - A Big Data Network Approach to Patent Valuation, Virginia Journal of Law and Technology, Vol. 20, No. 3, pp. 466-504.

Patents cite previous references relevant to their claims. These are known as backward citations. In turn, patents are cited by newer patents if the former are relevant to claims in the latter. These are known as forward citations. Both backward and forward citations can provide useful information about (1) a patent's value or importance and (2) about where the technology disclosed in the patent is situated within the wider universe of technological fields.

Citations to and from patents tend to be indicators of both private value to their owners and social value to society more generally.[65] Patent citations have been widely used in patent valuation analysis.[66] They can be rich sources of information about firm value[67], useful in assisting universities to predict which of the patents they own will most likely be licensed[68], and indicative of whether a patent application will be granted.[69] Patent citations have been found to correlate well with likelihood of litigation.[70] In fact, both backward and forward citations have also been found to be "unambiguously strong predictors of patent litigation", which has, itself, been found to be a robust indicator of high patent value.[71]

[65] Trajtenberg, M. 1990. A penny for your quotes: patent citation and the value of innovations. *Rand J. Econ.* 21:172-187.
[66] Bronwyn H. Hall, Adam B. Jaffe, & Manuel Trajtenberg, Market Value and Patent Citations: A First Look, National Bureau of Economic Research (2000); Dietmar Harhoff, Francis Narin, F.M. Scherer, & Katrin Vopel, Citation Frequency and the Value of Patented Inventions, 81 Rev. of Econ. & Stat. 511, 511 (1999); Lanjouw, J.O., & Schankerman, M. 2001. Characteristics of patent litigation: a window on competition, *Rand J. Econ.* 32:129; Adam B. Jaffe & Gaétan de Rassenfosse, *Patent Citation Data in Social Science Research: Overview and Best Practices*, 68(6) Journal of the Association for Information Science and Technology 1360 (2017).
[67] Hall, B.H., Jaffe, A.B., and Trajtenberg, M. 2005. Market value and patent citations. *Rand J. Econ.* 36:16-38.
[68] Sampat, B., & Ziedonis, A. 2004. Patent citations and the economic value of patents. In *Handbook of Quantitative Science and Technology Research*, ed. Moed, H.F., Glänzel, W., and Schmoch, U., pp. 277-298. Dordrecht: Kluwer Acad.
[69] Palangkaraya, A., Webster, E., and Jensen, P. 2011. Misclassification between patent offices: Evidence from a matched sample of patent applications. *Review of Economics and Statistics* 93(3): 1063-1075.
[70] Lanjouw, J.O., & Schankerman, M. 2002. An empirical analysis of enforcement of patent rights in the United States, Working Paper, p.4.
[71] John R. Allison et al., Valuable Patents, 92 Geo. L.J. 435, 451 (2004).

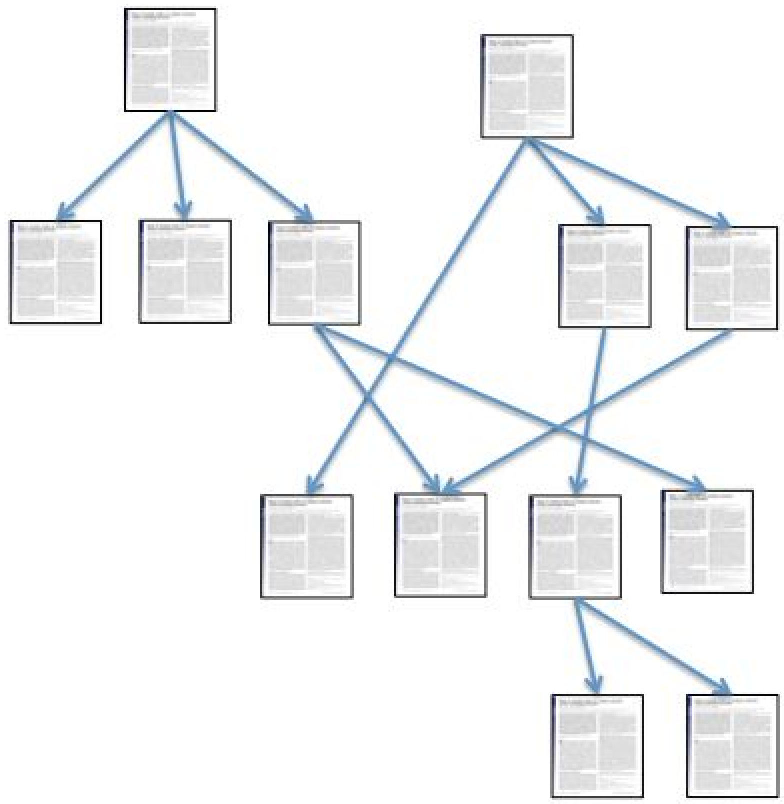
Beyond economic value alone, forward citations can provide good estimates of the technological importance of inventions disclosed in patents.[72]

Citation analysis of the scholarly literature also has a rich history, resulting in the standalone fields of bibliometrics and scientometrics. Librarians initially used citations to make journal subscription decisions.[73] This led to measures of journal prestige,[74] article quality,[75] author influence,[76] and even national intellectual output.[77]

A. PATENT CITATION NETWORKS

More than half a century ago, De Solla Price noted the utility and structural properties of citation networks.[78] In patent citation networks, the nodes represent patents and the links represent citations between patents and the non-patent literature. A simple schematic of a patent citation network is shown in Figure 2.

[72] Carpenter, M., Narin, F., and Woolf, P. 1981. Citation rates to technologically important patents. *World Patent Information* 3(4):160-163; Narin, F., Noma, E., and Perry, R. 1987. Patents as indicators of corporate technological strength. *Research Policy* 16(2-4):143–155; Albert, M., Avery, D., Narin, F., and McAllister, P. 1991. Direct validation of citation counts as indicators of industrially important patents. *Research Policy* 20(3):251-259.
[73] Gross, P., & Gross, E. (1927). College libraries and chemical education. Science, 66(1713), 385–389.
[74] Garfield, E. (1972). Citation analysis as a tool in journal evaluation. Science, 178(4060), 471–479.
[75] Walker, D., Xie, H., Yan, K., & Maslov, S. (2007). Ranking scientific publications using a model of network traffic. Journal of Statistical Mechanics: Theory and Experiment, 6, P06010.
[76] Hirsch, J. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences, 102(46), 16569–16572; West, Jevin D., et al. "Author-level Eigenfactor metrics: Evaluating the influence of authors, institutions, and countries within the social science research network community." *Journal of the American Society for Information Science and Technology* 64.4 (2013): 787-801.
[77] May, R. (1997). The scientific wealth of nations. Science, 275(5301), 793.
[78] de Solla Price, D. (1965). Networks of scientific papers. Science, 149, 510–515.



B.

Figure 2: Patent Citation Network. Nodes in this network are patents and the links are citations. This type of network is time-directed in that random walks on these citations go inexorably backwards in time. This schematic contains 13 nodes and 12 links. Our network contains more than 6 million nodes and more than 60 million nodes.

All patent citations are not equally useful as indicators. A citation by a patent's listed inventor to her own previous work (*i.e.*, self-citation) would probably merit different weight than a citation to the same patent by a scientist highly-influential in the patent's technological field. Many past studies involving patent citation data have relied upon raw citation counts. A more powerful way to appropriately weight citations is to construct a patent citation network in which the positions of each patent helps determine its value. Citation networks represent hundreds of millions of decisions by scholars that can help bibliometricians trace the influence of ideas and inventions.

There have been many proposed metrics for extracting the structural information from citation networks. One of the authors of this paper developed the Eigenfactor[79] metrics, which have been the gold standard in ranking scholarly journals. They are now included in Thomson-Reuters' Journal Citation Reports (JCR). The underlying algorithm is similar to the PageRank algorithm[80] developed by the founders of Google, Larry Page and Sergey Brin. The algorithm captures a random walker following hyperlinks (links) from webpage (nodes) to webpage. The Eigenfactor algorithm captures a random process on scholarly citation networks. For patent citation networks, we use a modified version of the Eigenfactor algorithm called the article-level Eigenfactor (ALEF).[81] The algorithm placed 1st in North America and 2nd worldwide in Microsoft Research's WSDM Cup Challenge, a 2015 contest whose goal was to statically rank tens of millions of articles from the scholarly literature.[82] The contest provided additional evidence of the advantages of using the network rather than just counting raw citations. To calculate the ALEF scores, we constructed a comprehensive patent citation network that includes all issued U.S. patents from 1976 to 2015 using methods described by West and Vilhena[83] and West *et al*[84].

Most patent citation networks calculated in the past have been national in scope. Recently, as detailed international bibliographic has become more readily available, networks of regional (e.g., members of the EPO) or even global scope have become more feasible. This allows comparisons to be made between national and international patent

---

[79] Jevin D. West, Theodore C. Bergstrom & Carl T. Bergstrom, *The Eigenfactor Metrics: A Network Approach to Assessing Scholarly Journals*, 71 C. & Res. Libr. 236 (2010).
[80] Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project. Retrieved from http://ilpubs.stanford.edu:8090/422/
[81] West et al (2016). Ranking and mapping article-level citation networks. In prep.
[82] Microsoft Research. 2015. WSDM Cup Challenge. Main website: https://wsdmcupchallenge.azurewebsites.net/; Winners leaderboard: https://wsdmcupchallenge.azurewebsites.net/Home/Leaderboard.
[83] West, J.D., and Vilhena, D. 2014. A network approach to scholarly evaluation. In *Bibliometrics & Beyond: Metrics-Based Evaluation*, ed. Cronin, B., and Sugimoto, C.R., pp. 151-166. Cambridge: MIT Press.
[84] West, J.D., Torrance, A.W., Rosvall, M., Vilhena, D., and Bergstrom, C.T. 2013. Systems and methods for data analysis. PCT Application (Filed February 1, 2013).

citations networks.[85] Global patent citations networks have been used to investigate particular areas of technology.[86]

IV. MATERIALS AND METHODS[87]

[85] Greg Morrison, Eleftherios Giovanis, Massimo Riccaboni, and Fabio Pammolli (2013), *Global and domestic centrality in patent citation networks* (located at: http://knowescape.org/wp-content/uploads/2013/11/knowescape2013_submission_20.pdf.) (This is a brief 2-page overview that aims to quantitatively determine high-impact central patent classes that are likely to provide domestic spillovers of info. It employs patent class citation networks, assuming info generated by government funding predominantly benefits a single class, and describes the motion of original information moving about the patent citation network as a random walk to allow them to determine a measure of centrality of each patent class. The authors suggest there may be differences in the importance of patent at the global versus national level. They find that, in all cases there is an overall decrease in the measured centrality as the variable increases due to the increase teleportation probability at every step of the random walk as that variable increases. They explain that the relative decline when comparing two central patent classes is due entirely to the topology of the citation network, and is a signal of the difference between global and domestic centrality. The U.S.' highly ranked classes were robust to variations compared to more rapid drops for other countries due to the overall global centrality of the U.S. economy. In Japan and Germany, as the variable increases, there is a greater degree of rearrangement, with the most central patent class globally not being the most central domestically. They conclude that, for these types of countries, national investment strategies may benefit by not only considering not only the global network topology but also by incorporating the domestically-centered measure of centrality. The method involved introducing an additional bias against the information crossing a political border, where there was no consequence of border crossing, and where information is effectively destroyed by border-crossing. They then tuned between global and domestic measures of centrality to determine the patent classes that are of primary domestic benefit to each individual country. They provided the results from the top five globally most central patent technology classes for the U.S., Japan, and Germany.)

[86] Hochull Choe, Duk Hee Lee, Il Won Seo, Hee Dae Kim. (2013.) Patent citation network analysis for the domain of organic photovoltaic cells: Country, institution, and technology field, 26 Renew. Sust. Energ. Rev., 492. (This study investigated worldwide patents in an attempt to understand the structure and characteristics of technological knowledge flows between countries, institutions, and tech fields in the field of organic photovoltaic cells. Using network topological analysis, network visualization, and node centrality, the article found that citation networks in this technological field are scale-free, follow the power law, and display a more efficient knowledge transfer capability than a random network of the same size. Node centrality analysis indicated that patents from the U.S,, Japan, and Germany are the most important citation centers in the network, and, of all USPTO technology classes, classes 136, 257, and 428 possess the most import core nodes. The authors also found that results from network topological analysis and node centrality analysis are not significantly different.)

[87] Parts of this section are adapted from Torrance, A.W., and West, J.D.. (2017) All Patents Great and Small - A Big Data Network Approach to Patent Valuation, Virginia Journal of Law and Technology, Vol. 20, No. 3, pp. 466-504.

In this section, we explain in detail the methods, databases, and analyses used to explore relationships between patents litigated to a decision in federal courts and the separately-derived importance of those patents.

Our data starts with the PATSTAT database[88]. This includes the bibliographic metadata (titles, publication date, inventors, citations, etc.) for about 100 million patent documents. It is publicly available for purchase from the European Patent Office. We extracted and transferred the data into a MySQL database. We then run software that we have developed to construct the citation network. The citation networks are stored as Pajek (network) files. These files provide information about the nodes (patents) and the links (citations). These network files are then used as input for our ranking and clustering algorithms. These rankings and algorithms have been described in previous publications[89]. We have developed software for identifying patent communities and automatically labeling the technology areas.

A. DATA SOURCES

We used the spring 2017 PATSTAT bibliographic database as the basis for our analyses. This database included detailed information on citations among patents and patent applications, examination and granting countries, priority, grant, and issue dates, inventor names, owners and assignees, USPTO, International Patent Classification ("IPC"), and Cooperative Patent Classification ("CPC") technology classifications, titles, and abstracts. There are approximately 46,000,000 granted utility patents[90], 800,000 design patents, 20,000 plant patents, and 16,000 reissue patents. There are about an additional 57,000,000 patent applications. The dataset includes patent documents from 174 jurisdictions and institutions, including the World Intellectual Property Organization

[88] https://www.epo.org/searching-for-patents/business/patstat.html#tab-1
[89] Adapted from Torrance, A.W., and West, J.D.. (2017) All Patents Great and Small - A Big Data Network Approach to Patent Valuation, Virginia Journal of Law and Technology, Vol. 20, No. 3, pp. 466-504
[90] These patents are known as utility patents in the United States, because the U.S. also grants design, plant, and reissue patents, but are often simply referred to as "patents" in other jurisdictions.

(WIPO), which is responsible for Patent Cooperation Treaty ("PCT") applications, the European Patent Office, which issues European Patents, and the African Regional Intellectual Property Organization ("ARIPO"). These patents and patent applications are connected by approximately 225,000,000 citations.

B. PATENT CITATION NETWORK

We previously assembled the largest patent citation network known in the literature, which was published as *All Patents Great and Small - A Big Data Network Approach to Valuation*. as of the writing of this Article. It includes 130 million citations (*i.e.*, "*links*") from nearly 6 million US patents (*i.e.*, "*core nodes*") from 1976 to 2015. The core nodes include about 4.5 million utility patents, 450k design patents, 20k plant patents, 16k reissue patents, 2k statutory invention registrations, and 500 defensive publications. About half the citations from these core nodes point to other core nodes (66 million citations). The other 60 million citations point to another 20 million nodes, which includes non-US patents from other countries, to patents from before 1976, and to non-patent references[91]. Citations are also labeled as originating from the inventor or the examiner. There were approximately 24 million examiner citations (15 million when isolated to core patent citations). For this analysis, we focus on the core nodes and the citations (both inventor and examiner citations) to/from the core nodes. This resulted in about 6 million nodes and 60 million citations.

Most patents receive a small number of citations, but there are some patents that receive a large number of citations. The highest cited patent in our database is U.S. Patent Number 4,683,202 ("Process for amplifying nucleic acid sequences"), invented by Kary Mullis, with more than 3000 citations.[92] The average degree (the number of in-citations to

---

[91] Non-patents citations are citations to items like the scholarly literature, books, newspapers, manuals, websites, etc. There are approximately 10 million citations.

[92] Unsurprisingly, this patent claims the polymerase chain reaction ("PCR"), which is one of the foundational technologies underlying biotechnology. The New York Times characterized this invention as "virtually dividing biology in the two epochs of before P.C.R. and after P.C.R.." Kary Mullis shared the 1993 Nobel Prize in Chemistry for this invention. (http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1993/) (Website last visited January 3, 2016).

each patent) is 10.3 in-citations per patent. This large number of citations per patent creates a dense network, with high in-degree and out-degree. When compared to other citations networks like the scholarly literature, this is highly dense for a citation network, and probably reflects the affirmative legal obligation under U.S. patent law for patent applicants to cite relevant and material prior art to the USPTO. Remarkably, there are fewer than 30k completely isolated patents that have neither backward nor forward citations.

Our database includes patent number, patent application number, patent title, USPTO technology classification codes, IPC technology classification codes, CPC technology classification codes, assignees, inventors, and both citations by each patent or patent application. The citation network allows for more complex queries using these data attributes. For example, one can identify the emergence of technology fields and the influencers of these technologies of fields using citations over time. Since the database includes patents from the PATSTAT database that span many years, any of these fields can be queried either statically (*i.e.*, at a particular point in time) or dynamically over any included range of times. In addition to the data fields derived from the patent metadata itself, the patent citation network can be analyzed using real-time, 'natural' technology clusters, which are groups of otherwise-unrelated patents that have strong mutual affinities within the network. To determine these natural clusters, we use the MapEquation framework.[93] We have compared USPTO, IPC, CPC, and our natural classifications of technology groups. Since patents within natural clusters generated in the patent citation network are strongly related to one another, we find that the accuracy of these natural clusters is higher than groupings based on USPTO, IPC, or CPC classifications.

V. RESULTS

A. DESCRIPTION OF THE WORLDWIDE PATENT CITATION NETWORK

The patent network consists of nodes (patents) and links (citations between patents). In this worldwide network, there are 103,096,180 nodes. This includes patents and

[93] Rosvall, Martin, Daniel Axelsson, and Carl T. Bergstrom. "The map equation." The European Physical Journal Special Topics 178.1 (2010): 13-23.

applications from all countries, regional jurisdictions, and institutions represented in the PATSTAT database. There are 46,486,933 granted patents in our dataset; these are the network nodes. The dataset also includes more than 223,307,509 citations. The average in-degree, which is the number of citations into each patent, is 6.6 citations per patent. This is lower than for the strictly U.S. patent network we previously constructed[94], which had an in-degree of 10.3, but 6.6 remains very high compared to most other kinds of citation networks. This decrease is expected given the looser connections to other countries versus connections to patents within the same jurisdiction. The average out-degree, which is the average number of references per patents, also differs. The average out-degree for the U.S. patent network alone is 21.0, whereas the world network has an average out-degree of 8.1 references per patent. This means that patents from some countries have considerably fewer references per patent than do patents from the U.S.. One explanation for this could that patents from countries that issue fewer patents are also cited fewer times.

The a disproportionate number of patents and citations come from the U.S., but many other countries are represented. In the future, we plan to calculate the ratio of in-country citations to out-country citations. This ratio will be useful in understanding the weight of the cross-country citations. These citations will play a special role in mediating the flow of ideas (via citations) across the network. This is a very promising method for providing clues to technology trends and flows among countries. Patents may originate in large multinational companies that pursue patent coverage in multiple countries, but they may also be have especially high value. Our methods are effective at measuring these flows, not just for individual patents, but also for millions or tens of millions of patents. Our methods allow measurement of these inflows and outflows among countries over time.

We map the citation network using a modified version of InfoMap, which is a software platform for clustering large networks. These clusters provide information about what patents are grouped together, using only the citation networks. This information can be used to supplement the traditional technology categorization tags (*e.g.*, IPC). The difference is that our mappings can be used in real time, and often differ from

[94] Torrance, A.W., and West, J.D.. (2017) All Patents Great and Small - A Big Data Network Approach to Patent Valuation, Virginia Journal of Law and Technology, Vol. 20, No. 3, pp. 466-504.

categorizations based on human judgment. We can also identify far more fields that can be classified using traditional tagging and classification techniques. With our analysis, we find 10,7115,34 distinct technology subfields. The average size per technology field is 77.35 patents. This indicates a highly skewed distribution of fields, with some fields having a lot of patents, but most fields having relatively few. The clustering is done hierarchically, so the smaller fields aggregate up to bigger fields. Going from only U.S. patents to the worldwide patent network led to more fields, but the distribution of cluster sizes per field has remained similar to that of the U.S. patent citation network.

B. COUNTRIES BY PV SCORE

The worldwide patent citation network is composed of approximately 130,000,000 patents and patent applications. Figure 3 shows how these patent documents are distributed among national jurisdictions and WIPO.

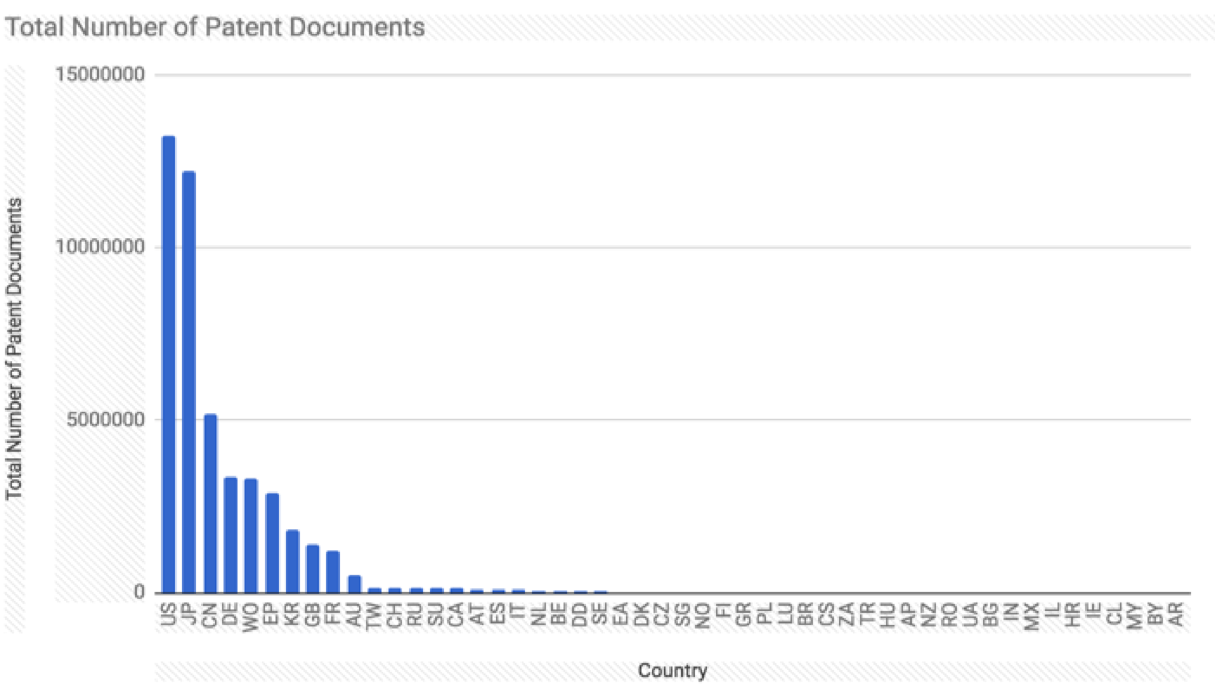


Figure 3. Total number of patents and patent applications by country or organization.

The U.S. has the most patent documents, with about 13,000,000. Japan is close behind with about 12,000,000. After these two countries, there is a steep decline to the third-place country, China, which has just over 5,000,000 patent documents. After another substantial decline in numbers, both Germany and WIPO are next with about 3,500,000 each, followed by the EPO with about 3,000,000. With another substantial drop patent documents, South Korea is next with about 2,000,000, then the United Kingdom and France with about 1,500,000 each. Australia is next with about 500,000. After additional sharp drop in numbers of patent documents come Taiwan, Switzerland, Russia, Sweden, the former Soviet Union, and Canada, each with about 100,000.

Raw numbers of patents and patent applications fail to convey the importance or value of patent rights generated within the patent systems of countries or organizations. To accomplish this, we calculated the PV Score of every patent document. Figure 4 shows the total aggregate importance of patents and patent applications as distributed among national jurisdictions and WIPO.

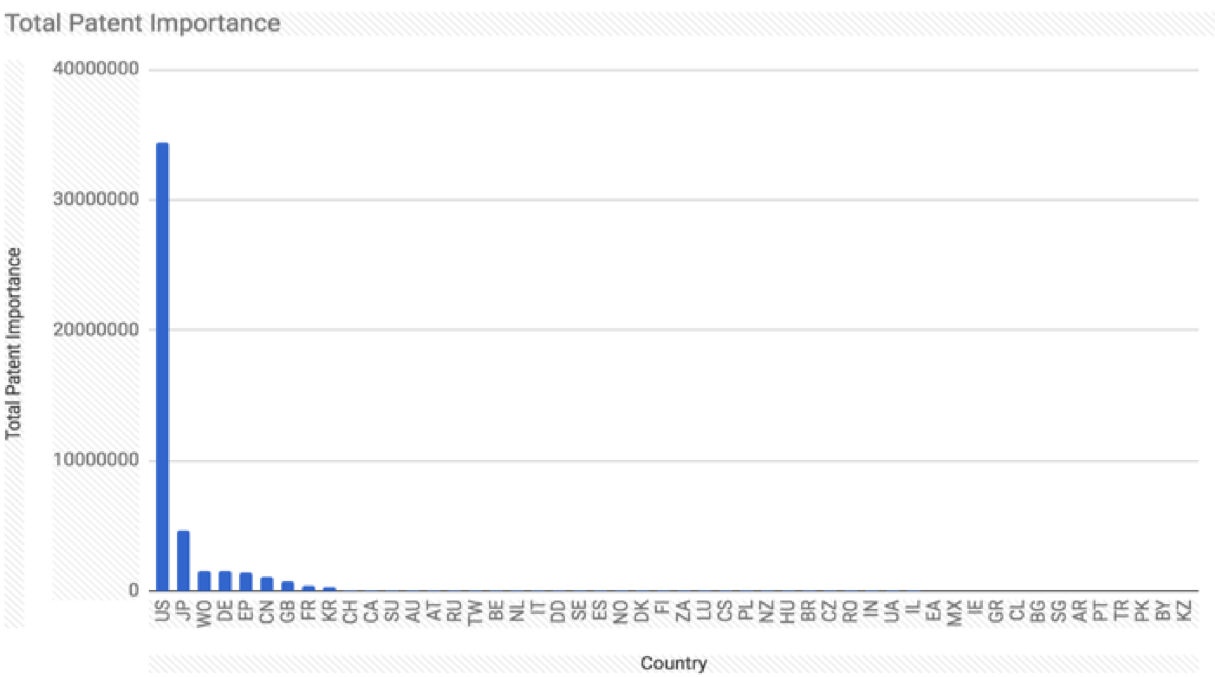


Figure 4. Total aggregate importance of patents and patent applications by country or organization.

When importance or value is taken into account, by calculating and adding up the PV Scores of all patent documents, the international league table of patents and patent applications changes substantially. The U.S. has the highest total patent importance, with an aggregate PV Score of about 35,000,000. No other country or organization is close to this level of patent importance. Japan comes second with an aggregate PV Score of about 4,000,000, or roughly one tenth the patent importance of the U.S.. The WIPO, Germany, and EPO each have aggregate patent importances of about 1,500,000. China is next with a total patent importance of about 1,000,000. The United Kingdom follows with an aggregate patent importance of about 500,000. France is next, with a total patent importance of about 200,000. South Korea has an aggregate patent importance of about 100,000. After that, no other country of organization is remotely close in aggregate patent importance.

We also calculated the mean PV Score of patents and patent applications by country or organization. Figure 5 shows the mean importance of patents and patent applications as distributed among national jurisdictions, WIPO, and the EPO.

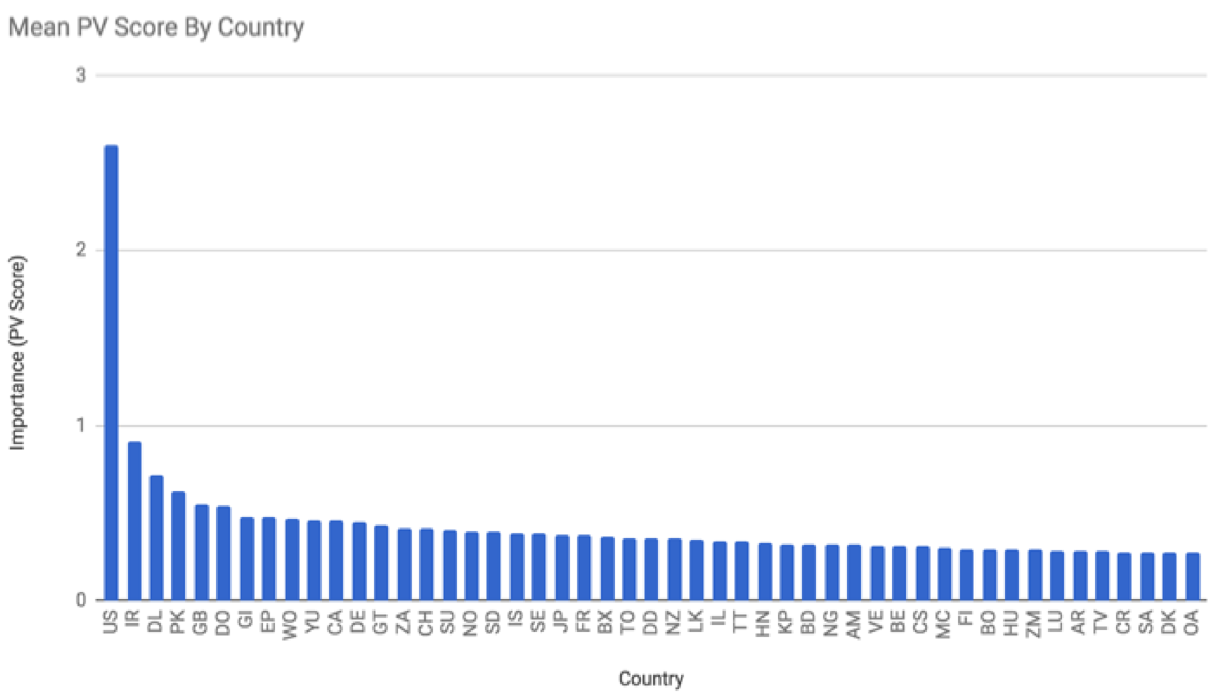


Figure 5. Mean importance of patents and patent applications by country or organization.

Mean patent importance tells a somewhat different story, with the differences among the top countries and organizations much less extreme than it is for aggregate patent importance. Again, the U.S. leads in importance, with a mean PV Score of about 2.6. Iran places second in patent importance with a mean PV Score of about 0.9; however, this is based on only four patents. Next is the defunct German Democratic Republic, whose mean PV Score of about 0.7 is also based on a very small number of patents. Pakistan is next, with a mean PV Score of about 0.6. The United Kingdom and the Dominican Republic are next, having mean PV Scores of about 0.5.[95] Gibraltar, the EPO, the WIPO, the former Yugoslavia (including only Serbia and Macedonia), Canada, and Germany each have mean PV Scores of about 0.4. The next 38 countries in Figure 5 have mean PV Scores ranging from less than 0.4 to about 0.3.

C. ASSIGNEES BY PV SCORE

Figure 6 depicts which patent assignees have the largest worldwide patent portfolios.

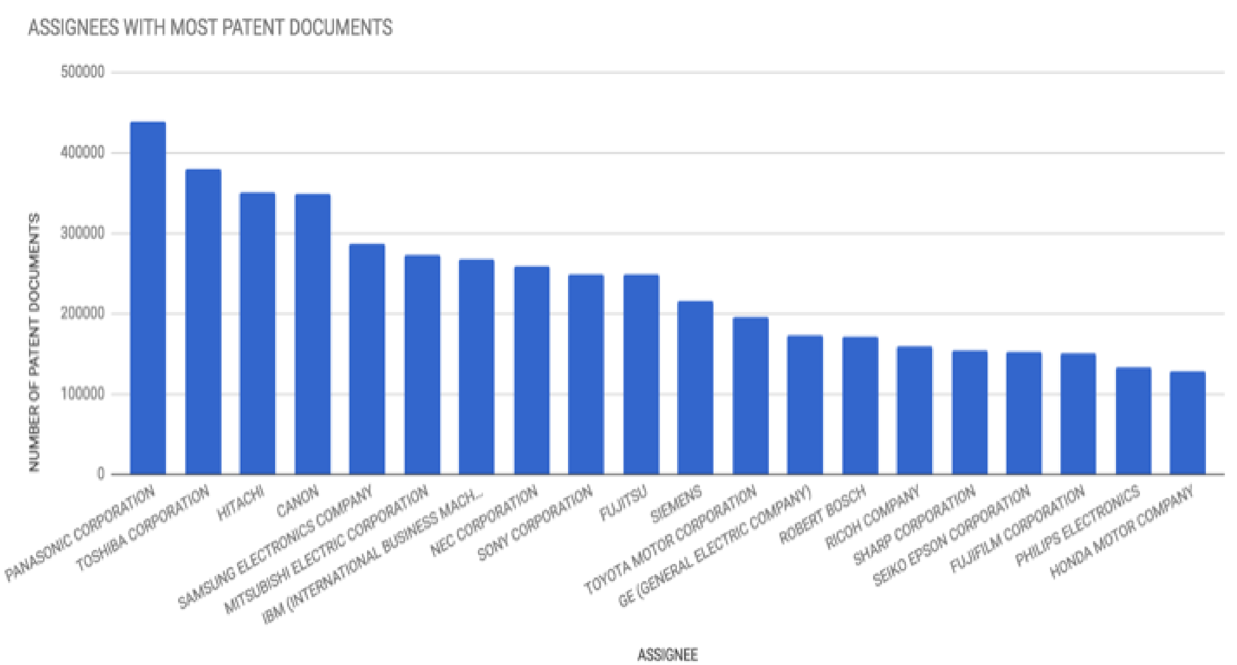


Figure 6. Assignees with the largest patent portfolios.

[95] It is worth noting that the size of the U.K. patent corpus is many times larger than that of the Dominican Republic.

The largest of these is Panasonic Corporation, with about 440,000 in total. The four largest patent portfolios, as well as eight of the largest ten, belong to Japanese companies. Only Samsung Electronics Company, a South Korean firm, and International Business Machines, a U.S. company prevent a clean sweep of the top ten. Filling out the rest of the top 20 largest patent portfolios are six more Japanese firms, two German companies (*i.e.*, Siemens and Robert Bosch), one Dutch firm (i.e., Philips Electronics), and one U.S. company (*i.e.*, General Electric Company).

Figure 7 depicts which patent assignees have patent portfolios having the highest aggregate importance.

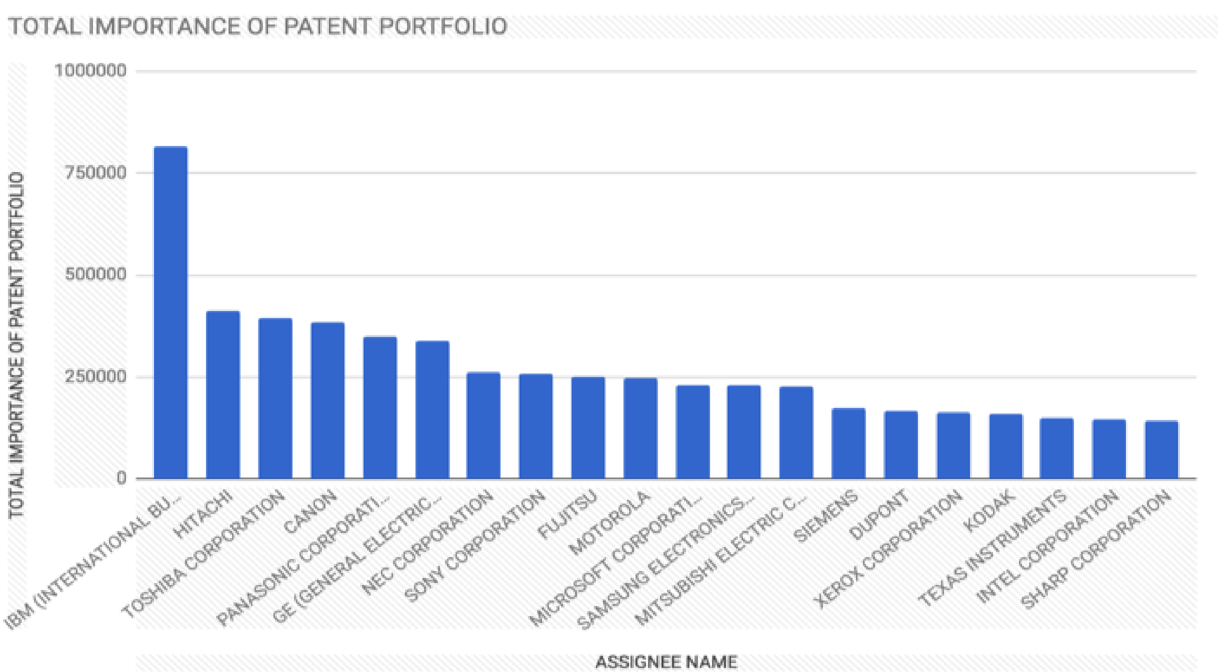


Figure 7. Assignees with the most important patent portfolios.

Although many of the companies in this top 20 list are also found in the top 20 list of largest patent portfolios, when portfolios are measured by importance their relative order changes markedly. International Business Machines possesses the most important patent portfolio by a wide margin, with double of the aggregate importance of the second place company, Hitachi. In fact, U.S. companies account for 9 of the 20 most important patent portfolios. Japanese companies are also well represented, with 9 companies in the top 20. The top 20 also contains one company each from Germany and South Korea.

Figure 8 shows which patent assignees have patent portfolios whose constituent patents have the greatest mean importance.

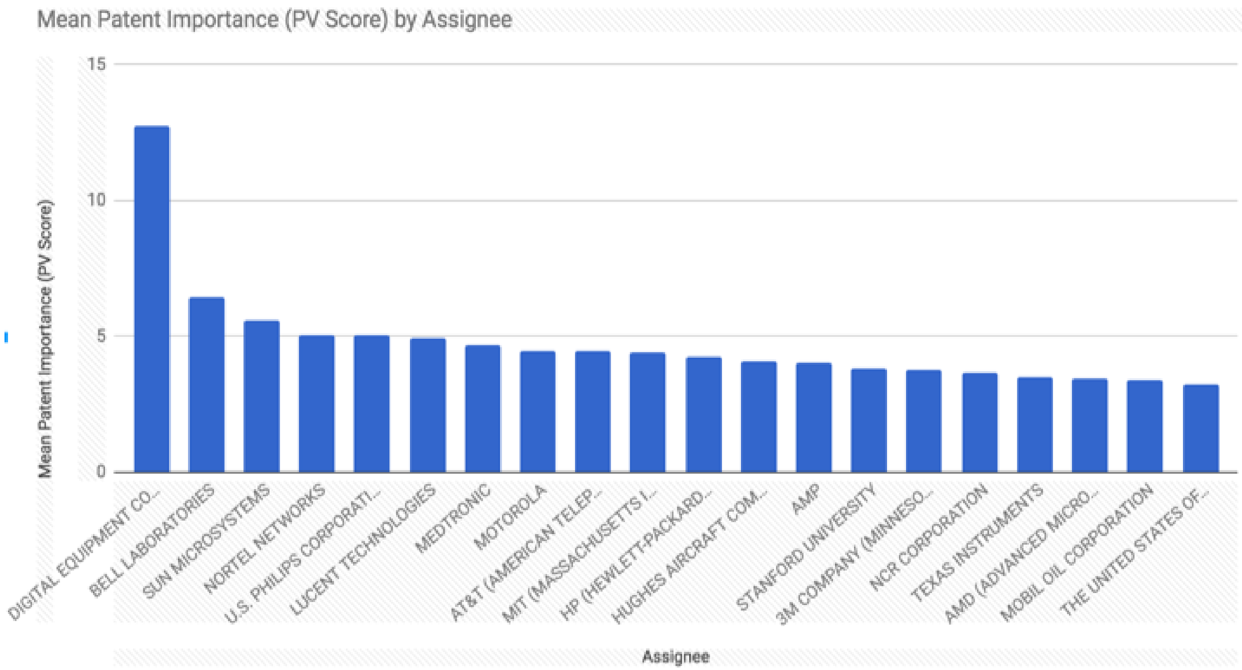


Figure 8. Assignees with portfolios of the highest mean PV Score.

When mean patent importance is taken into account the composition of the top 20 assignees changes substantially from that of the top 20 lists for size of portfolio and aggregate importance. Digital Equipment Corporation possesses the highest mean patent importance, which is double the importance of the mean patent in the portfolio of the second place company, Bell Laboratories. There are two universities (*i.e.*, the Massachusetts Institute of Technology and Stanford University) in the top 20, as well as the United States federal government. Finally, only one assignee in the top 20 is not from the U.S.: Nortel Networks, a Canadian company that was dissolved in bankruptcy in 2009.

D. INVENTORS BY PV SCORE

Figure 9 shows the top 30 inventors as measured by mean patent importance.

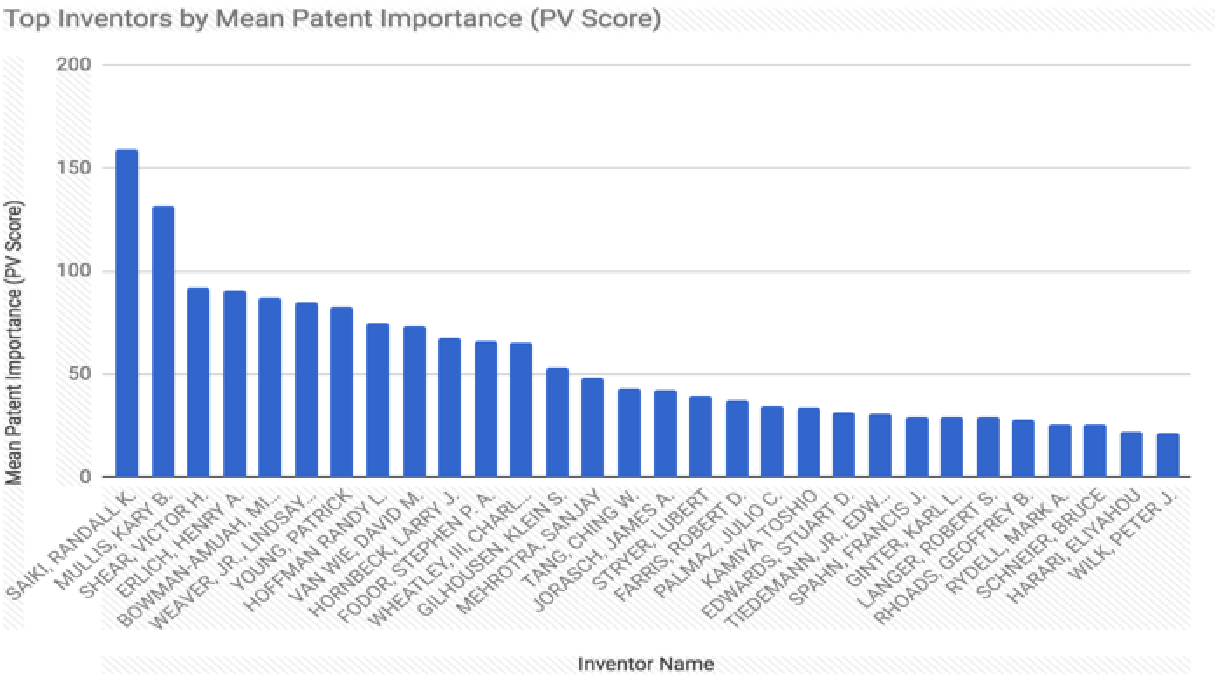


Figure 9. Inventors of patents and patent applications having the highest mean PV Score.

Almost all of these inventors either hold (or held) U.S. citizenship or are (or were) based in the U.S..

E. MOST IMPORTANT INVENTIONS

Table 1. shows the top 14 most important patents worldwide.[96]

Table 1. The most important patents worldwide.

[96] Importance in Tables 1 and 2 is indicated with a U.S. dollar value calculated as follows: we multiply the mean inflation-corrected value of a U.S. patent, as estimated by Bessen, J. (2006) (The Value of Patents by Owner and Patent Characteristics, Boston University School of Law Working Paper Series, Working Paper No. 06-46) by the PV Score of that patent. This estimate of patent value does not take into account expiration, invalidation, or unenforceability of individual patents.

US	4683202	Process for amplifying nucleic acid sequences	$360,466,830.46
US	4683195	Process for amplifying, detecting, and/or-cloning nucleic acid sequences	$326,933,946.63
US	4733665	Expandable intraluminal graft and method and apparatus for implanting an expandable intraluminal graft	$216,793,350.83
US	5572643	Web browser with dynamic display of information objects during linking	$210,843,513.91
US	4816567	Recombinant immunoglobin preparations	$201,752,570.23
US	4799156	Interactive market management system	$187,278,094.43
US	5143854	LARGE SCALE PHOTOLITHOGRAPHIC SOLID PHASE SYNTHESIS OF POLYPEPTIDES AND RECEPTOR BINDING SCREENING THEREOF	$152,652,969.43
US	4503569	Transluminally placed expandable graft prosthesis	$141,704,084.61
US	4655771	Prosthesis comprising an expansible or contractile tubular body	$141,418,647.68
US	5892900	Systems and methods for secure transaction management and electronic rights protection	$139,587,436.32
US	5220501	METHOD AND SYSTEM FOR REMOTE DELIVERY OF RETAIL BANKING SERVICES	$138,207,165.67
US	5523520	Mutant dwarfism gene of petunia	$132,913,568.43
US	4580568	Percutaneous endovascular stent and method for insertion thereof	$131,004,135.71
US	4395486	Method for the direct analysis of sickle cell anemia	$129,485,724.20

These patents range widely across technical categories, including molecular biology, medical devices, and software.

Table 2. shows the top 14 most important patents granted in Canada.

Table 2. The most important Canadian patents.

publn_auth	publn_nr	appln_title	pv_value
CA	971285	DYNAMIC ADDRESS TRANSLATION REVERSED	$421,755.45
CA	1072413	POLOXAMER GEL SYSTEMS WITH GELLING TEMPERATURES HIGHER THAN ROOM TEMPERATURE	$401,403.33
CA	1162336	SYSTEM FOR ENTERING AND PROCESSING GAMBLE STAKES	$379,770.23
CA	2015507	SPINAL IMPLANT	$335,086.74
CA	2012311	NEW PROTEIN-POLYCATION-CONJUGATE	$328,981.44
CA	2111317	SYSTEM, METHOD AND APPARATUS FOR GENERATING LARGE JACKPOTS ON LIVE GAME CARD TABLES	$320,974.06
CA	1044639	POLYMER-BOUND PHOTOSENSITIZING CATALYSTS AND PHOTOSENSITIZED REACTIONS UTILIZING SAME	$304,325.78
CA	737249	CHECK VALVE SYSTEM	$303,908.64
CA	911410	PHOSPHATE SUBSTITUTED ZEOLITE COMPOSITIONS AND METHOD FOR PREPARING SAME	$302,693.71
CA	1086178	LIQUID DETERGENT COMPOSITION	$295,457.78
CA	2036946	INDEXING LINKERS	$292,505.01
CA	2088442	METHOD AND SYSTEM FOR FILE SYSTEM MANAGEMENT USING A FLASH-ERASABLE, PROGRAMMABLE, READ-ONLY MEMORY	$290,481.05
CA	803714	CONTINUOUS FILAMENT FABRIC	$284,527.70
CA	649852	PHOTOTROPIC LIGHT FILTERS	$281,459.52

These patents have far lower importances than do those in the global importance list. They also range across a variety of technological areas, including chemistry, molecular biology, medical devices, fabric, and gambling systems.

F. TECHNOLOGIES BY PV SCORE

Figure 10 shows the top 30 CPC technology classifications by aggregate patent importance.

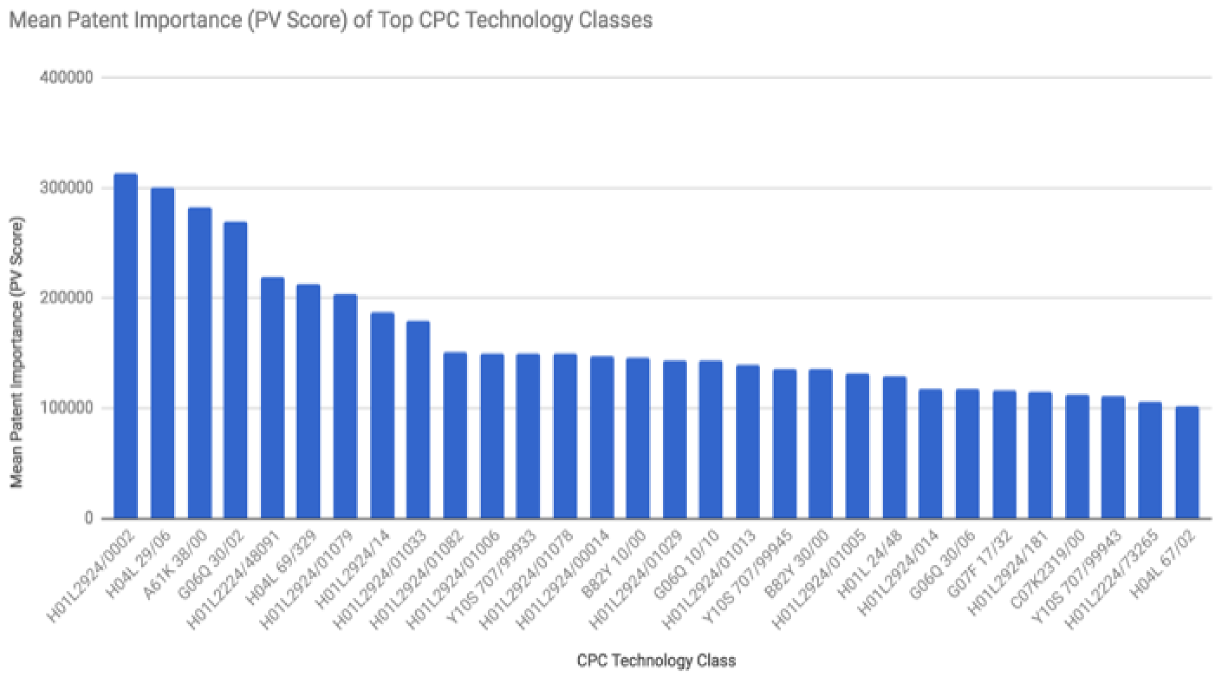


Figure 10. Technology classes (CPC) with the highest aggregate PV Score.

Of the top 30 CPC technology categories, a large majority (*i.e.*, 19 out of 30) belong to classification H, which covers inventions involving electricity (*e.g.*, computers, mobile phones).

Figure 11 shows the top 30 CPC technology classifications by mean patent importance.

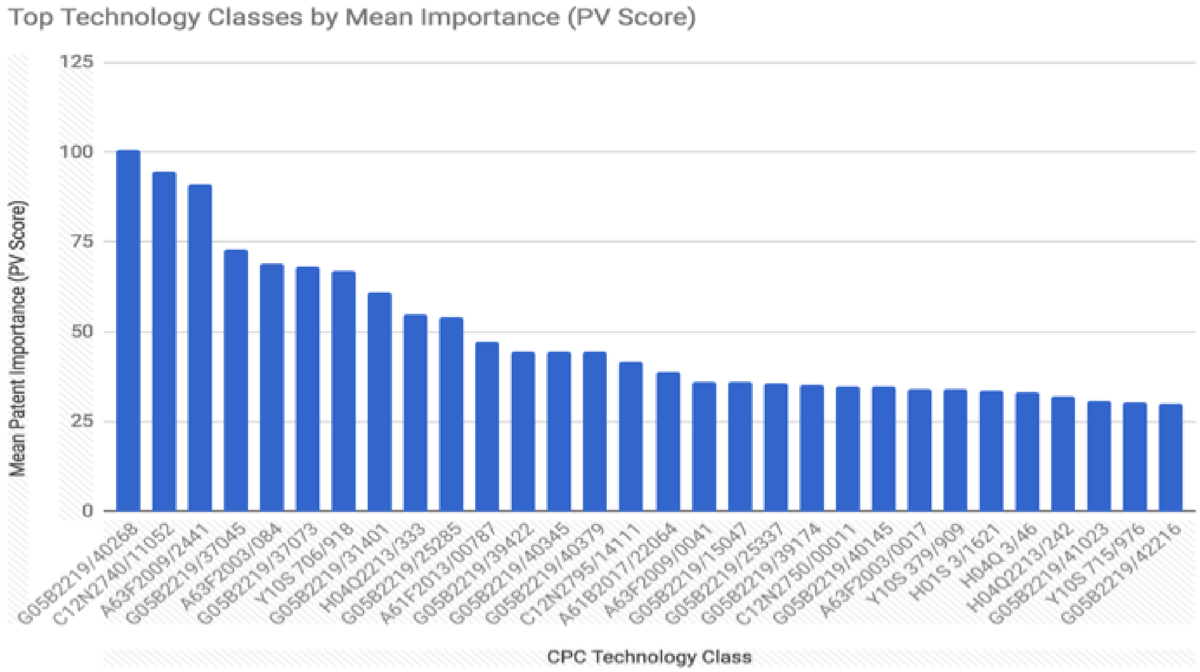


Figure 11. Technology classes (CPC) with the highest mean PV Score.

Of the top 30 CPC technology classifications, only four are in classification H. Almost half (*i.e.*, 14) belong to classification G (*i.e.*, physics), 6 to classification A (*i.e.*, human necessities), 3 to C (*i.e.*, chemistry; metallurgy), and 3 to Y (*i.e.*, general miscellany).

VI. DISCUSSION

A. MORE DATA IS BETTER

One adage in data science is that more data means better data. However, this is not always the case. In the case of the worldwide patent citation network we have calculated, the larger citation network reveals insights not only about the relative influence of patents of countries around the globe, but it also reveals new information about the U.S. patent system. When ranking and clustering the patent citation network, citations to U.S. patents from non-U.S. patents provides another layer of information about the U.S. patents that is unavailable when only examining the U.S. citation network in isolation. We can also construct patent citation networks without using any U.S. citations. Given the importance of U.S. patents and the citations they generate, we predict that such a network would be much less complex, and much more disconnected, than the worldwide patent citation is. Nevertheless, a global network that excluded U.S. patents may provide interesting insights about the patent influencers in the rest of the world.

The most important piece of information that non-U.S. citations can provide is the relationship between patents. If patent A and patent B from the U.S. are cited by the same patent in another country, that can indicate that patent A and B are related in some meaningful way not obvious when only considering the U.S. patent citation network. When we scale this to tens of millions of citations, as we have done with the worldwide patent network presented here, it is possible to reveal new communities of patents that are lost with the U.S.-centric view.

B. TRACING INNOVATION FLOWS ACROSS COUNTRIES OVER TIME

One of the most exciting areas of analysis that can be one when viewing the patent system as a large citation network is the ability to trace flows of information across fields, countries, and regions. This can highlight companies that cite U.S. patents, technology fields trending upwards or downwards in activity and influence, inventors origins, and the effect patents have on collaboration across countries. We are currently constructing maps of innovation flow across countries, though this process is time-consuming and computationally challenging .

C. GENERAL OBSERVATIONS

Constructing a worldwide patent citation network is a very challenging computational problem. The building of the patent citation "tree" itself requires a powerful computer and currently takes more than a week of continual calculation.

A few features of the worldwide patent citation network are particularly worthy of note. Almost all of the most important patents worldwide are U.S. patents. In fact, 99 of the 100 most important are U.S. in origin, with the single exception a European (EP) patent. The mean U.S. patent is quite important, with a mean PV Score of 2.60. The next large economy in terms of patent importance is the United Kingdom, which has a mean PV Score of 0.547, followed by Canada with a mean PV Score of 0.459. High mean patent importance is not simply a function of inventing in English; Australian patents have a mean PV Score of 0.095. By way of comparison, China and India both have much less important patents, with mean PV Scores of 0.217. It is also noteworthy that there is considerable variation in patent importance based on inventor, assignee, and technology class, both within and among countries and institutions.

D. FUTURE RESEARCH

We plan a number of additions and improvements on this study. These include adding the full texts of patent abstracts to our analysis. This will require some form of reliable automated translation because many abstracts are unavailable in English. Eventually, we will add full specification texts as well, but this presents substantial computational challenges. In addition to the text of specifications, we plan to add figures and tables, though this too is a formidable computations challenge.

There is a wealth of information beyond what is directly part of patent documents. Once high fidelity data about such valuable information as litigation, ownership entity type, standards-essential patents are available, we hope to add these to our analyses. Eventually, we hope to create a zoomable graph of the worldwide patent citation network analogous to Google Earth, to facilitate visual exploration of the network, its citations, and its nodes. Such a visual graph may also allow viewing of a worldwide “heatmap” that indicates technological areas in which much or little activity is present. Eventually, we also hope to reconstruct past worldwide patent citation networks from various defined points in time to allow better analysis of the evolution of technologies.