Journal of International & Interdisciplinary Business Research

Volume 3 *Journal of International & Interdisciplinary Business Research* Article 5

January 2016

CALCULATING THE EUCLIDEAN TECHNOLOGY DISTANCE OF DYADS USING PATENT CITATIONS

Don Goeltz
Holy Family University, dgoeltz@holyfamily.edu

Jan Buzydlowski
Holy Family University, jbuzydlowski@holyfamily.edu

CALCULATING THE EUCLIDEAN TECHNOLOGY DISTANCE OF DYADS USING PATENT CITATIONS

Don Goeltz, Holy Family University
Jan Buzydlowski, Holy Family University

Calculating the technology base of a firm is a critical first step in studies of the technology strategies of a single entity and in making comparisons between the technology strategies of firms. For example, many studies of alliances and alliance portfolios require calculation of the technology distance or technology overlap between firm dyads. These studies typically use the patents of each partner dyad as the bases for the calculation. This paper introduces a new method of calculating a measure of technology overlap using the patent citations made in the patent applications. Each patent application lists the patented technologies that are being cited, much like citations in an academic paper. Use of various distance calculations using patents and the use of patent citations to evaluate technology direction are accepted concepts in the literature. This paper combines and advances these ideas by using patent citations in the distance calculation. By examining patent citations, we calculate technology distance of a dyad at a broader and deeper level than would be available by looking only at the patents themselves. The paper first examines the current techniques for calculating technology overlap and summarizes some of the current applications. Then, the new calculation technique is derived and examples of the calculation are presented. Finally, areas for further research are explored.

CALCULATING THE EUCLIDEAN TECHNOLOGY DISTANCE OF DYADS USING PATENT CITATIONS

Measuring technological distance is often a concern of managers and researchers when analyzing external alliance partners, mergers and acquisitions, and in developing an R&D strategy within a firm. Here we first define the terms technology distance, technology similarity, and technology overlap. Then the focus is on the applications of technology overlap and on current measurement approaches. A novel -- and arguably improved -- calculation technique based on patent citations is described and illustrated.

Technology distance can refer to firms, industries, geographic areas, and countries. In each case, a technology profile is calculated from either inputs, such as R&D spending or number of R&D researchers employed, or from outputs, such as patents or new product introductions. Technology distance can be decomposed into technologies that are complimentary to the focal entity and a measure of the technical overlap. This study focuses on the technology overlap measures and their applications.

INTRODUCTIONS AND DEFINITIONS

Researchers have addressed the effects of technology overlap in general and have also specifically addressed technology overlap in external learning. Although there are many potential dimensions to technology overlap, the majority of researchers use patent data as the measurement source and overlap of patent classes as the technology overlap measure.

For example, in looking at organizational learning, Bierly, Damanpour, and Santoro (2009) test the general association of technology overlap, as measured by patent portfolios, with performance in studying university R&D contracts, finding that less technology overlap was associated with increased exploratory innovation. Similarly, Quintana-Garcia and Benavides-Velasco (2008), in a longitudinal study of biotechnology firms, find support for the association of technology overlap with increased innovation. This study also finds that there are diminishing returns to technology overlap and that it had a stronger effect on exploration than on exploitation. Finally, Ahuja and Lampert (2001) find that exploration of a variety of novel, emerging, and pioneering technologies had an inverted U-shaped relationship with breakthrough inventions.

These results reflect that knowledge that is close to the existing base has advantages for learning in organizations. If the new knowledge is in the neighborhood of the existing knowledge base, then the firm can gain economies of scale by reusing the existing knowledge and gain from the learning curve effect (Hayward, 2002). The new knowledge will be easier to absorb as it shares codes, values, language, and symbols with existing knowledge (Grant, 1996). Similarity of product, customer, and managerial backgrounds facilitates synergies when the knowledge is related (Tanriverdi & Venkatraman, 2005). As a result, organizations often stay in the vicinity of their competencies, reinforcing existing routines (Cyert & March, 1992).

Some degree of knowledge overlap also helps the firm to recognize the value of new, external learning, to absorb it and to commercialize the value (Cohen & Levinthal, 1990). Therefore knowledge acquisition is enhanced when there is overlap between existing and new knowledge. However, if knowledge is too similar, not much new is acquired, there are fewer opportunities for innovative knowledge combinations, and few synergies occur. Learning by reusing existing knowledge is valuable for exploitation and progressing down a similar learning curve, but it is of much less value for external exploration investments where knowledge crosses organizational and technological boundaries (Rosenkopf & Nerkar, 2001).

The calculation of technology overlap is therefore an essential step in determining the value of an alliance partner or the value of a merger or acquisition. Calculation of the technology overlap of a firm's entire alliance portfolio is often referred to as the technology diversity of the portfolio. Firms also rely on this calculation when defining the long-term focus of an R&D project, with the objective that the new technology knowledge is somewhat related to the existing technology base.

EXISTING APPROACHES TO CALCULATING TECHNOLOGY OVERLAP

There are several ways to measure empirically the dimensions and positions of firms in a technology space. Broadly, there are patent-based measures and non-patent-based measures. With non-patent-based measures, the position in knowledge space can be determined by inventor or scientist characteristics (Adams, 1990; Farjoun, 1994). Another possibility is to analyze the R&D profile. For example, Goto and Suzuki (1989) measure the technological distance among 50 sectors based on the spending of R&D into 30 product areas. Sapienza, Parhankangas, and Autio (2004) use questionnaires asking Finnish CEOs of spin-offs about their assessment of the technological distance to their previous parent company.

Patent-based measures of technology overlap are primarily based on either counts of co-occurrences or on calculating distance in a multidimensional space. For example, technology overlap has been measured by the co-occurrence of classifications in patent documents (Breschi, Lissoni, & Malerba, 2003; Nesta & Saviotti, 2005; Schmidt-Ehmcke & Zloczysti, 2008).

The concept of a multidimensional space is most often applied to patents by using the patent classification system to define the dimensions. For each patent, the patent examiner determines a technology class based on the application. Angular separation (Jaffe, 1986) uses patent classifications to measure technological distance. It measures to what degree the vectors point in the same direction, controlling for the length of the vector by the number of patents in that dimension. Rosenkopf and Almeida (2003) propose the use of the Euclidean distance to compare the firms' technology vectors. This approach compares for each technology category the squared difference of the share of that technology category in the focal firm with the share that the technology class has in the object firm. Some authors subtract the Euclidean distance from one to obtain a measure that is decreasing in distance as overlap decreases. For an extensive review of the measures that can be used for determining overlap, the reader is referred to Stellner, 2014. Stellner (2014) makes it clear that some approaches are better suited for specific studies. The approach that is proposed here is based on studies of alliance dyads and alliance portfolios.

In a further refinement of the calculation of technology vectors, Sampson (2007) calculates the weighted Euclidean distance between two firms across the patent bases of the two firms. Construction of this variable starts with the generation of each partner's technological portfolio by measuring the distribution of its patents across patent classifications, year by year. This distribution is then captured by a multidimensional vector that represents the number of patents assigned to a firm in each patent class. Diversity of partner firm capabilities is then the weighted number of common occurrences of patents in each class.

The Sampson (2007) technique is the most frequently used measure of technological overlap in studying alliances, and it often results in an inverted U-shaped relationship between technological overlap and learning. Examples of use of the Sampson technique of calculating technology overlap include a study of partner selection in R&D alliances (Li, Hitt, & Ireland, 2008), the effects of technology alliances on innovation patterns (Zidorn & Wagner, 2013), and a study of acquisitions (Lee & Kim, 2014; Marki, Hitt, & Lane, 2010). This technique has also been used in examining the complementarity of partners (Noseleit & de Faria, 2013) and in balancing technology portfolios (Gilsing, Vanhaverbeke, & Pieters, 2014; Goeltz, 2014).

Researchers using the Sampson method for calculating technology overlap rely on classifications of the patent provided by the US Patent and Technology Office (USPTO), as described in the next section. The refinement of the Sampson calculation technique described here is based on patent citations, which are also products of the USPTO patenting process.

PATENT DATA AND THE USE OF THE PATENT SYSTEM

Patent data is arguably the best source of information for the measurement of technological distance, given its fine-grained split of technological categories. Patent data is readily availability from the USPTO, the World Intellectual Patent Office, and the International Patent Classification (IPC) system. The USPTO patent has two data fields that are important for calculating technology overlap – the technology class and the patents that are cited as foundation for the new patent. The IPC has similar fields.

The USPC (US Patent Classification) organizes all U.S. patent documents into a class and a subclass. A class generally delineates one technology from another. Subclasses delineate processes, structural features, and functional features of the subject matter encompassed within the scope of a class.

A USPC classification uniquely identifies one of the over 400 classes and more than 150,000 subclasses. A complete identification of a subclass requires both the class and subclass number and any alpha or decimal designations; e.g., 417/161.1A identifies Class 417, Subclass 161.1A. Every U.S. patent document has at least one mandatory classification, and may optionally include one or more discretionary classifications. The principal mandatory classification is known as an OR classification and can have any number of secondary classifications for additional claims in the patent application.

Not only the applicants of the patent add patent citations, but also by the examiners of the patent application. Patents citations are determined by the examiner who, with the help of the data supplied by the applicants and their attorney, determines whether specific citations are relevant or not (Leydesdorff & Fritsch, 2006). The number of citations a patent has can also be linked to the market value of the company owning the patent and the value of the technology (Hall, Jaffe, & Traijtenberg, 2005).

In 2014 there were 615,234 patent applications to the USPTO and 326,033 patents granted. The elapsed time between application and grant is, on average, 32 months. Given that applicants can repeatedly apply and adjust their applications, the actual patent approval rate in 2012 was almost 90% (USPTO).

The USTPO maintains a searchable database of patents from 1790 to the present. Associated with each patent are the primary classification and any number of secondary classifications. Also associated with each patent are the prior patents cited as a basis for the new patent. However, as this data is not easily compiled into reports, a large number of services exist to do custom searches and reports.

Problems with the USTPO Patent Database

Given the extent of the USTPO patent database, it is the source that is used most often in research into technology overlap. However, there are a number of problems that arise from the use of USTPO patent data (Griliches, 1990). A few problems relevant to this research topic are:

a) The propensity to patent depends largely on the size of the company, given the long time between submission and grant and the expense of documentation.
b) The propensity to patent varies significantly among technologies. Hence, technologies with a lower propensity to patent are underrepresented (Jaffe, 1986; Stellner, 2014).
c) The technology profile of companies and the resulting distance measure depend on a single classification given to each patent, typically a four-digit class/subclass identified at the primary classification. In some cases, researchers use higher and lower levels of classification, with little justification for the choice (Stellner, 2014)
d) The classification of patents is subject to errors on the part of patent examiners and there are potential biases, in that patent examiners may classify patents in fields with which they are familiar (Stellner, 2014).

One way to address these issues is to broaden the data that is captured by using patent citations as a data source. Patent citations appear in the Cited Patent Section of each patent. While the patentee may have the incentive to cite as few other patents as possible in a patent application, a patent examiner verifies the correctness of patents cited and eventually demands that other patents be cited before granting (Lanjouw & Schankerman, 1999).

When the U.S. Patent and Trademark Office grants a patent, the granting officer includes a list of all previous patents on which the granted patent is based. Citations of prior patents thus serve as an indicator of the technological lineage of new patents, much as bibliographic citations indicate the intellectual lineage of academic research (Mowery, 1996).

In a study of why inventors reference patents, the researchers find that 57% of the citations provide a foundation of knowledge spillover, and 43% the patent attorney or patent examiner add the citation (IEEE, 2010). Nonetheless, the IEEE study recommends using all available data points in studies. A study of the bias introduced by using just the primary patent classification, it was recommended that that researchers increase their sample size (Benner & Waldfogel, 2008).

Prior efforts to utilize the patent citations include citation counts and cross-citation rate. For example, in a study of acquisitions, Seats and Hoetker (2013) calculate overlap by dividing the count of common citations and patents by the total count of patents and patent citations of the acquirer and target firms. A method, described below, was developed to take advantage of the patent citation data and to address the shortcoming of current approaches that use the single data point of the primary technology classification assigned to each patent.

THE PATENT CITATION APPROACH TO MEASURING TECHNOLOGY OVERLAP

As mentioned, each submitted patent is classified according to the International Patent Classification (IPC) system. This provides only a single classification for each patent; however, as also mentioned, other patents which are related to, or support, the submitted patent are included in the Cited Patent Section of that patent. Using the classification of those patents in addition to the single classification provides more information. Intuitively, it makes sense that the supporting and similar patent classifications would be related to, and expansive of, the submitted patent.

For example, if Company A is submitting (or was granted) Patent 3, then the technology categories of the cited patents (say, Patent 1 and Patent 2) are used to construct a vector of technology categories in addition to the single classification of Patent 3. This will be denoted as a LAG-1 vector. In this case it will be three classification codes, one from each of the patents involved. It is also of interest to look back for a second connection; i.e., to look at the supporting patents for each supporting patent of the original patent. This will be designated as a LAG-2 vector. For example, the cited patents for Patents 3, Patents 1 and 2, are then each searched for their cited patents, say Patents 5, 6 for Patent 1 and Patents 7 and 8 for Patent 2, and those technology categories are also used. This will result in a collection, or vector, of seven codes. Naturally, actual patents will contain many more than just two supporting patents, so the numbers in actuality will be significantly greater.

Having arrived at a method and implementation of obtaining a technology vector for a patent, the next step is to determine the overlap in technology categories between two companies; i.e., the dyad. The process starts with measuring the distribution of a firm’s patents across patent classifications, year by year. This distribution is then captured by a multidimensional vector that represents the number of patents assigned to a firm in each patent class. Diversity of partner firm capabilities is then the weighted number of common occurrences of patents in each class. The technologic vectors are derived from the patents that were granted for those companies within a defined time window of the alliance. The two technology vectors from the dyad are then calculated and stored.

As the final step, the determination of the amount of overlap between the two vectors in the alliance period, the statistic described in Sampson, 2007 is used:

$$1 - \frac{F_i F'_j}{\sqrt{(F_i F'_i)\,(F_j F'_j)}},$$

where F_i is a vector containing the counts of the technology categories associated with a patent for company i, F_j is a vector containing the counts of the technology categories associated with a patent for company j, and F'_i F'_j are the transpose of the corresponding vectors. A value near 1 indicates complete technology diversity, i.e., there is no overlap between the two entities. Values near 0 indicate complete overlap.

This citation approach of measuring technology overlap using an expanded category collection offers advantages over existing approaches as it broadens the information base. It will also more accurately calculate the technology overlap variable for companies that patent less often, typically smaller or more secretive companies.

Applications of the Patent Citation Approach with Examples

To illustrate the above theory to compute the technologic overlap between company dyads, an analysis based on actual patent data is illustrated below.

The National Bureau of Economic Research (NBER) has embarked on a project to make the patent citations of all patents available in an online database; consequently, the patent citation data for this analysis is from the National Bureau of Economic Research website (Roth, 2015).

From that dataset that NBER makes available, the two datasets of interest are the Patent data, including constructed variables (pat63_99.txt) and the Pairwise-Citations data (Cite75_99.txt). The first file contains the patents awarded from the years 1963 to 1999 and contains numerous other fields, such as geographic location as well as technologic categories and subcategories. For a full description of the accompanying data, the reader is referred to the documentation linked to in (Roth, 2015). The latter file, Cite75_99.txt, contains the cited patents for the patent requested.

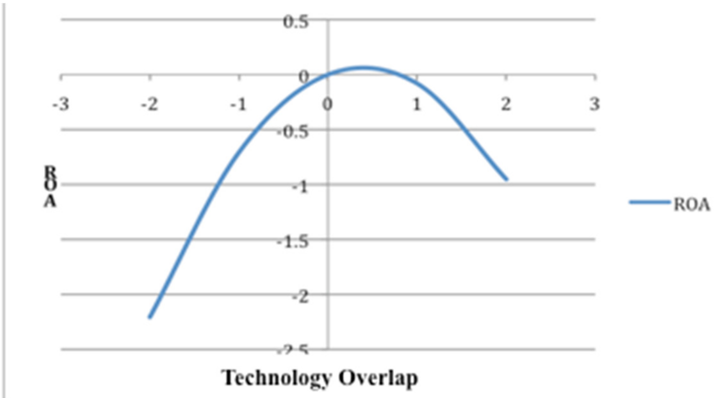
The datasets are sizable. For the Patent data, there are 2,993,922 company records for the granted patents, which represented the patents granted to companies between the years 1963 and 1999. Each record has 23 fields consisting of elements such as company ID, application and grant year (10 fields), as well as the categories and subcategories of the grant (13 fields). For the Pairwise-Citation data, there are 16,522,439 records with two fields, citing patent number and cited patent number. It is through this second file that the technology overlap is obtained. Whereas, if looking exclusively at the first file, only one category was applied to that patent; however, by looking at the cited patents, i.e., those patents that are cited in the application of the original patent to justify that patent, and then by referring back to the original technology patent and obtaining that category for the cited patent, a list of technology categories is built based on the original patent.

To support this computation, a relational database is used and queries (SQL) are built to find the LAG-1 and LAG-2 vectors. As the data made available via the website (Roth, 2015) is simple CSV-formatted data, population of the database is relatively easy as most databases allow the simple importing of such file formats. For the Patent data, the data is formatted in the data tables as expected, e.g., county as text, but traditionally textual data is converted to numeric fields for speed and space considerations. For this set, the primary keys are the patent number as well as the only indexed variable. For the Pairwise-Citation data, there are only two fields, but since this dataset is constantly used in the SQL for the lag computations, and, as indicated above, quite sizable, both fields are stored as large integers and both are indexed to increase efficiency.

To do the computations, a simple database, Microsoft Access 2010, on a standard operating system, Windows 7, on a common hardware platform, a laptop with 4G RAM, Intel Pentium running at 1.3 GHz, and 400 GB HD is used. This platform was found to be sufficient for the purposes of this analysis.

To compute the overlap using the Sampson Statistic, the vectors were extracted from the database and exported to a CSV file. From there, an implementation of the Samson Statistic is implemented in Python. The implementation is done in a little over 100 lines of Python code and included modules to read in the data, read in the companies of interest from an Excel file, and export the data to another Excel file, as well as the computation of the Sampson Statistic. (Python is chosen as it is a scripting language, so the code needed is small, and works well with other programs, such as Excel, Access, or reading/writing any file format.)
Performing the above steps of finding the technology vectors and computing the similarity measures resulted in very interesting results. For example, as predicted, the technology diversity has an inverted-U relationship with firm performance (Figure 1).

Figure 1. Regression of Technology Overlap with Return on Assets



From the dataset of this study, we illustrate further uses of the measure, for example how the technology overlap differs between firms in the same industry. Within the telecommunications industry the technology overlap of Cisco (Figure 2) and Motorola (Figure 3) have different profiles over time, with Cisco lower but rising, while the technology overlap of Motorola with its alliance partners is decreasing over time.

Figure 2. Technology Overlap of Cisco with Alliance Partners

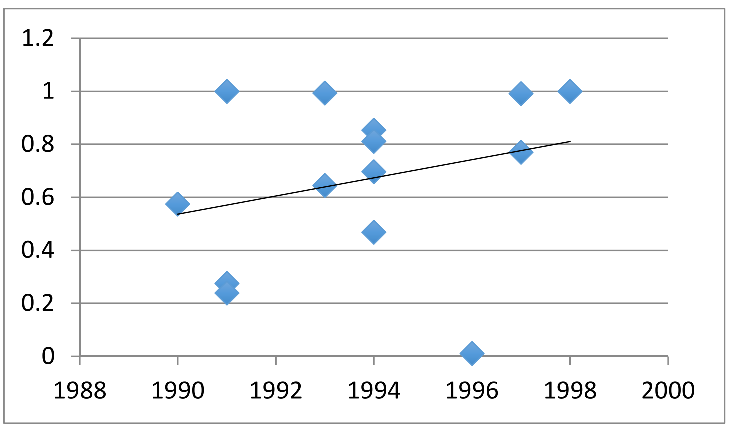
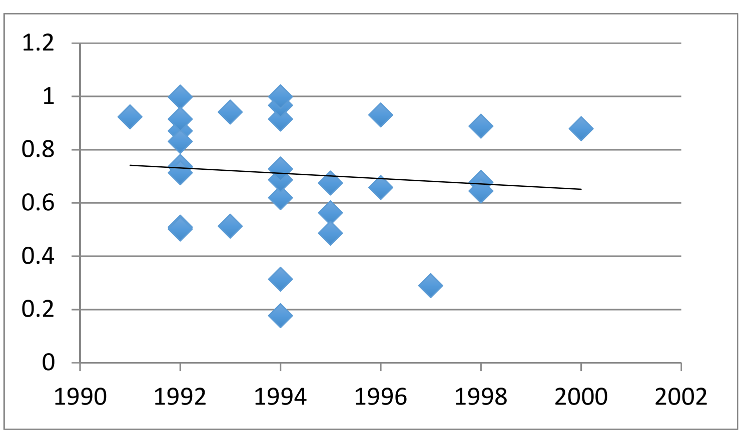


Figure 3. Technology Overlap of Motorola with Alliance Partners



This comparison raises several relevant research questions concerning these trends at the firm level and their relationship to industry, competitive, and financial results.

Another level of analysis is opened up by comparison of the technology overlap of firms with the industry average and trends. For example, in the biotech industry Enzon's technology overlap is trending toward one (Figure 4) while the overall industry trend is well below that level (Figure 5).

Figure 4. Technology Overlap of Enzon with Alliance Partners

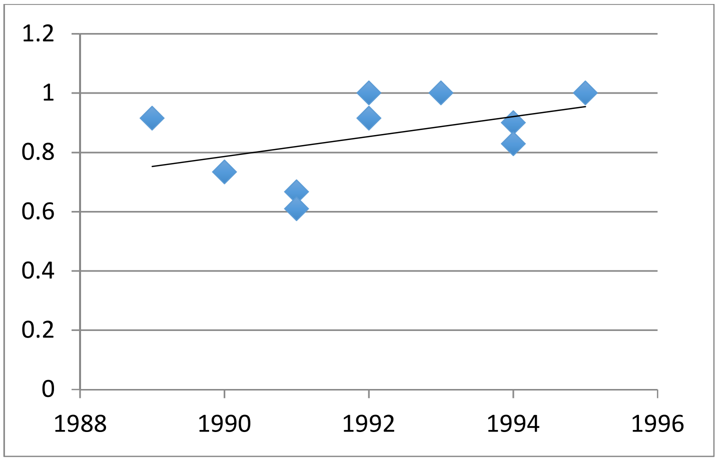
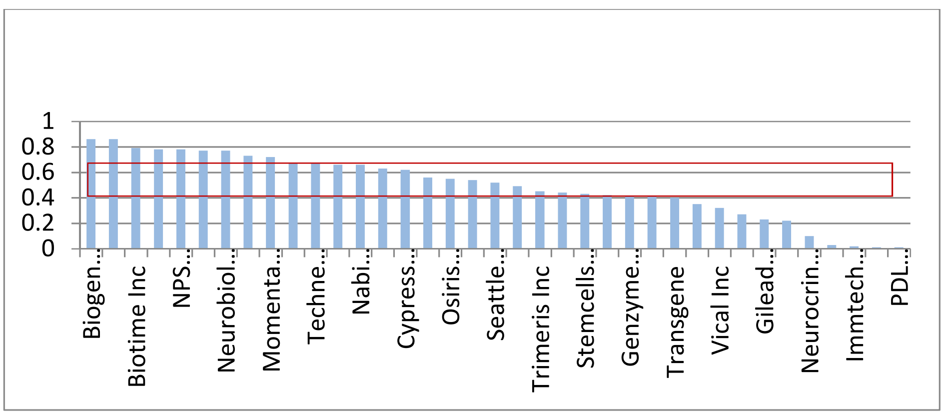


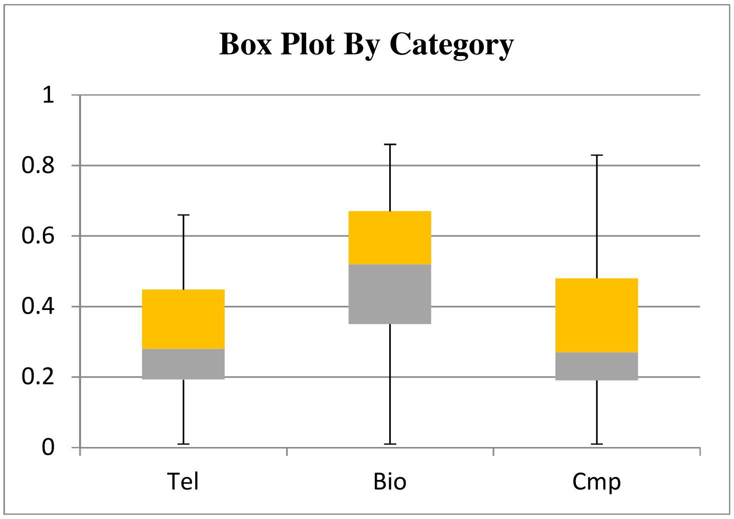
Figure 5. Biotech Industry Technology Diversity



Again, this type of analysis opens up further research questions, such as how the level of a firm's portfolio within an industry could be a leading or a lagging indicator of performance.

In the final example, the comparison of technology diversity at the intra-industry level can also be explored. As shown in Figure 6, the measurements among the biotech, telecommunications, and computer services industries are quite different (whisker ends indicate extreme values).

Figure 6. Comparison of Industry-Level Technology Diversity



LIMITATIONS AND AREAS FOR FUTURE RESEARCH

There are several limitations within the patent citation approach to calculating technology overlap and its applications. First, this paper and approach is based on the study of alliance portfolios. It may or may not be applicable to other research areas. Most of the other limitations are the same ones that apply to the use of the primary patent classification. First, the classification itself is determined by a patent author and is verified by a single patent examiner. Second, big companies patent more frequently and more broadly than small firms, making comparisons problematic, although this is mitigated somewhat by using the broader base of citations. Finally, this analysis technique is based on the NBER dataset, which currently lags the patent dataset by ten years, limiting the timespan of research.

Areas for future research are examined in the prior section at the intra-firm, inter-firm, industry, and country levels. At the firm level, one potential application is to track the technology portfolio over time. If a firm's technologies exhibit a great deal of overlap, that could mean that the company should expand its R&D scope and/or expand the boundaries of the technology space through alliances and acquisitions. A study in this area could relate the firm's technology overlap to market and financial performance. Likewise the direction of the trend line, as illustrated in Figures 3 and 4, could be studied as a leading indicator of the level of competitiveness of a firm, which is useful in competitive strategy studies.

At the firm and inter-firm levels, the trend in technology overlap can be studied and compared to the innovation rates of companies as measured by patents issued and new product/service introductions. Also at the inter-firm levels, technology overlap can be compared to financial performance measure over time to see if diversification leads to improved performance.

Further study at the industry and firm-level analysis is illustrated in Figure 5, comparing a firm's technology overlap to an industry average. This analysis could be done as a time series, again looking at the relationship to innovation and financial performance measures.

Figure 6 illustrates a way of comparing industries. Industry-level analyses could include how the technology diversity of the industry changes over time, the influence of global competition on the variable, and a comparison of the R&D intensity at the industry level with the industry technology diversity measure.

The nature of this calculation is that there is no accepted standard across all applications. However, the authors have considered a side-by-side comparison of the patent citation approach with the single-measure patent primary technology classification approach to investigate whether a study of alliance portfolios using the alternative approaches.

In terms of potential evolution of the patent citation approach, the authors are considering ways of combining the patent citation technique with co-occurrence analysis as well as information retrieval methodologies to further broaden the base of the calculation. They are also looking at ways of examining the various technology overlap calculations to find which technique is best suited for different types of analysis.

The authors welcome comments on and corrections to these application areas, as well as on the development of the patent citation technique. The authors would also like to than the reviewers of the paper for their constructive comments and additions.

REFERENCES

Adams, J. D. (1990). Fundamental Stocks of Knowledge and Productivity Growth. Journal of Political Economy, 98 (41), 673-702.

Ahuja, G., & Lampert, C. (2001). Entrepreneurship In The Large Corporation: A Longitudinal Study Of How Established Firms Create Breakthrough Inventions. Strategic Management Journal, 22, 521-543.

Benner, M., & Waldfogel, J. (2008). *Close to you? Bias and precision in patent-based measures of technological proximity. Research Policy*, *37*(9), 1556-1567.

Bierly, P., Damanpour, F., & Santoro, M. (2009). The Applications Of External Knowledge: Organizational Conditions For Exploration And Exploitation. Journal Of Management Studies, 46(3), 481-509.

Breschi, S., Lissoni, F., & Malerba, F. (2003). Knowledge-Relatedness In Firm Technological Diversification. Research Policy, 32, 69-87.

Cohen, W., & Levinthal, D. (1990). Absorptive Capacity: A New Perspective On Learning And Innovation. Administrative Science Quarterly, 35, 128-152.

Cyert, R. M., & March, J. G. (1963). A Behavioral Theory Of The Firm. Englewood Cliffs, NJ: Prentice Hall.

Farjoun, M. (1998). The independent and joint effects of the skill and physical bases of relatedness in diversification. Strategic Management Journal, 19 (7), 611-630.

Gilsing, V. A., Vanhaverbeke, W. P. M., & Pieters, M. (2014) Mind the gag: Balancing alliance network and technology portfolios during periods of technological uncertainty. Technological Forecasting and Social Change. 81, p. 351-362.

Goeltz, D. R. (2014), Alliance portfolio characteristics and firm performance, Journal of Management and Market Research

Goto, A., & Suzuki, K. (1989), "R&D Capital, Rate of Return on R&D Investment, and Spillover of R&D in Japanese Manufacturing Industries," Review of Economics and Statistics, 11, 555-564.

Grant, R. M. (1996). Prospering In Dynamically-Competitive Environments: Organizational Capability As Knowledge Integration. Organization Science, 7, 375-387.

Griliches, Z. (1991) Patent statistics as economic indicators: a survey, Journal of Economic Literature 28, 1661-1707.

IEEE (2010). Why do Inventors Reference Papers and Patents in their Patent Applications? IEEE Publications

Hall, B. H., Jaffe, A. B., Trajtenberg M. (2000). Market Value and Patent Citations: A First Look. National Bureau of Economic Research: Cambridge, MA.

Hayward, M. (2002). When Do Firms Learn From Their Acquisition Experience? Evidence From 1990-1995. Strategic Management Journal, 23, 21-39.

Jaffe, A. B. (1986). Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market value. American Economic Review, 76 (5), 984 –1001.

Makri, M., Hitt, M. A., & Lane, P. J. (2010). Complementary technologies, knowledge relatedness, and invention outcomes in high technology mergers and acquisitions. Strategic Management Journal, 31(6), 602-628.

Leydesdorff, L., & Fritsch, M. (2006). Measuring the Knowledge Base of Regional Innovation Systems in Germany in terms of a Triple Helix Dynamics. Research Policy, 35(10), 1538-1553.

Li, D., Eden, L., Hitt, M. A., & Ireland, R. D. (2008). Friends, acquaintances, or strangers? Partner selection in R&D alliances. Academy of Management Journal, 51(2), 315-334.

Nesta, L., & Saviotti, P. (2005). Coherence of the Knowledge Base and the Firm's Innovative Performance: Evidence from the US Pharmaceutical Industry. Journal of Industrial Economics, 53(1), 123-142.

Noseleit, F., & de Faria, P. (2013). Complementarities of internal R&D and alliances with different partner types. Journal Of Business Research, 66(10), 2000-2006.

Python Foundation. (2015). Python Home. Retrieved from Python.org: https://www.python.org/

Quintana-García, C., & Benavides-Velasco, C. A. (2011). Knowledge organisation in R&D alliances: its impact on product innovation. Technology Analysis & Strategic Management, 23(10), 1047-1061.

Rosenkopf, L., & Nerkar, A. (2001). Beyond Local Search: Boundary-Spanning, Exploration, And Impact In The Optical Disk Industry. Strategic Management Journal, 22, 287-306.

Rosenkopf, L., & Almeida, P. (2003). Overcoming local search through alliances and mobility. Management Science, 49 (6), 751-766.

Roth, J. (2015). The NBER U.S. Patent Citations Data File: Lessons, Insights, and Methodological Tools. National Burea of Economic Research.

Sampson, R. C. (2007). R&D Alliances And Firm Performance: The Impact Of Technological Diversity And Alliance Organization On Innovation. Academy Of Management Journal, 50, 364-386.

Sapienza, H. J., Parhankangas, A., & Autio, E. (2004). Knowledge relatedness and post-spin-off growth. Journal of Business Venturing, 19 (6), 809-829.

Schmidt-Ehmcke, J., & Zloczysti, P. (2008). Technology Portfolio and Market Value (780). DIW Berlin, German Institute for Economic Research.

Stellner, F. (2014). Technological Distance Measures: Theoretical Foundation and Empirics. DRUID Society.

Tanriverdi, H., & Venkatraman, N. (2005). Knowledge Relatedness And The Performance Of Multibusiness Firms. Strategic Management Journal, 26, 97-119.

United States Patent and Technology Office. (USTPO) http://www.uspto.gov/web/offices/ac/ido/oeip/taf/us_stat.htm