Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Position and Differentiation of Firms in Technology Space

Sam Arts, Bruno Cassiman, Jianan Hou

informs
https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 69, No. 12, December 2023, pp. 7253–7265
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Position and Differentiation of Firms in Technology Space

Sam Arts,[a,*] Bruno Cassiman,[b,c] Jianan Hou[d]

[a] Department of Management, Strategy and Innovation, Faculty of Economics and Business, KU Leuven, 2000 Antwerp, Belgium; [b] Department of Management, Strategy and Innovation, Faculty of Economics and Business, KU Leuven, 3000 Leuven, Belgium; [c] Strategic Management Department, IESE Business School, University of Navarra, 08034 Barcelona, Spain; [d] Department of Business Management, School of Management, Xiamen University, 361000 Xiamen, China

*Corresponding author

Contact: sam.arts@kuleuven.be, https://orcid.org/0000-0003-3214-7325 (SA); bruno.cassiman@kuleuven.be, https://orcid.org/0000-0001-9602-6755 (BC); jianan.hou@xmu.edu.cn, https://orcid.org/0000-0003-3911-4948 (JH)

Received: December 14, 2021
Revised: June 17, 2023
Accepted: June 25, 2023
Published Online in Articles in Advance: September 21, 2023

https://doi.org/10.1287/mnsc.2023.00282

Abstract. Although the rate of invention by firms and the effect on firm performance have been central themes in economics and strategy, the position and differentiation of invention by firms have received less attention. We develop a method to characterize a firm's technology portfolio based on the semantic content of patents that allows us to map a firm's unique spatial position relative to every other firm in technology space and to measure the overall differentiation of a firm's technology portfolio. Using a large panel of U.S. public firms from 1980 to 2015, we illustrate that technology differentiation has a strong positive correlation with firm performance, particularly in research and development-intensive industries and industries with strong product market rivalry. We also show that technology differentiation is associated with subsequent differentiation from competitors in the product market and a reduction in outgoing technology spillovers to other firms. We provide open access to code and data to characterize the technology portfolio of firms and to measure the technological position and differentiation of U.S. public firms.

History: Accepted by Alfonso Gambardella, business strategy.

Funding: S. Arts acknowledges support from KU Leuven [Grant 3H200208] and National Bank Belgium [Grant 3H160338]. B. Cassiman acknowledges support from Fonds Wetenschappelijk Onderzoek - Vlaanderen [Grant G071417N].
Supplemental Material: The data files and online appendix are available at https://doi.org/10.1287/mnsc.2023.00282.

Keywords: technology differentiation • firm performance • technology portfolio • text analysis • patent portfolio

1. Introduction

The rate of invention by firms has been a central theme in economics and strategy. Prior studies have repeatedly shown that research and development (R&D) investments and the size of a firm's technology portfolio relate to superior firm performance (Hall et al. 2005, Simeth and Cincera 2016, Bellstam et al. 2021). In contrast to the rate of invention, the position and differentiation of invention by firms have received far less attention in the literature (Jaffe 1986, Stuart and Podolny 1996). Nevertheless, a firm's position and differentiation in technology space relative to other firms are arguably key drivers of firm performance in line with differentiation in product market space (Hotelling 1929, Shaked and Sutton 1982, Hoberg and Philips 2016, Guzman and Li 2023). According to the resource-based view, a firm's competitive advantage and superior performance relies on resources that are unique and difficult to substitute or imitate, such as a unique and proprietary technology portfolio (Wernerfelt 1984, Peteraf 1993, Mowery et al. 1998). Formal theory models suggest that firms tend to differentiate their technology from other firms to minimize outgoing technology spillovers and competition in the product market (Kamien and Zang 2000, Aghion et al. 2005, Gil Moltó et al. 2005, Lin and Zhou 2013).

To contribute to this stream of work, we develop a new method to characterize a firm's technology portfolio based on the text of patents that allows us to map each firm's unique spatial position relative to every other firm in technology space and to measure the overall differentiation of a firm's technology portfolio.[1] To characterize firm technology portfolios, prior work has mostly relied on the patent classification system of patent offices (e.g., Jaffe 1989, Rosenkopf and Almeida 2003, Bloom et al.

2013). However, patent classifications are perhaps too broad to capture the detailed content of firm technology (Thompson and Fox-Kean 2005, Arts et al. 2018, Righi and Simcoe 2019). Instead of relying on patent classifications, we exploit the fact that according to U.S. law, firms must provide a fully written disclosure of their inventions in clear and exact terms in exchange for legal patent protection.[2] Compared with the traditional approach based on patent classifications, patent text provides a more detailed insight into the technology portfolio of firms (Arts et al. 2018, Righi and Simcoe 2019), and it outperforms in the identification of new technologies (Arts et al. 2021, Bowen et al. 2023).[3] Identifying new technologies is important to continuously update the ever-changing technology space and to accurately measure a firm's position and differentiation in this technology space relative to other firms.

Our data collection draws from the Duke Innovation & Scientific Enterprises Research Network (DISCERN) database (Arora et al. 2021) and from the processed text of U.S. patents (Arts et al. 2021). Representing each firm-year-level technology portfolio as a vector based on the semantic content of patents, we calculate cosine similarities to measure the pairwise technology similarity between all U.S. public firms for all years and use these to construct our measure of technology differentiation for each firm and year. As expected, younger, smaller, and more R&D-intensive firms display higher levels of technology differentiation relative to older, larger, and less R&D-intensive firms.

Next, we study how our firm-year-level measure of technology differentiation predicts future firm performance. Although these estimates are not causal, they both provide a validation of our new measure and present new descriptive facts on the relation between technology differentiation and firm performance. Using firm fixed effects models on the entire panel of U.S. public firms for the years 1980–2015, we find that technology differentiation has a strong positive relation with market value. Controlling for a firm's R&D intensity and the number of citation-weighted patents in the technology portfolio, a one-standard deviation increase in technology differentiation corresponds with an increase of 10.4% in Tobin's *Q*. In addition, we find that, as expected, technology differentiation is particularly valuable in R&D-intensive industries and in industries with strong product market competition. Next, we empirically examine the two mechanisms advanced by theory models for why technology differentiation might increase firm performance. First, we show that technology differentiation is associated with subsequent differentiation from competitors in the product market as measured by the overlap in 10K business descriptions with industry rivals. Second, we demonstrate that an increase in technology differentiation is related to a decrease in outgoing technology spillovers to other firms. Finally, we show that when technology differentiation is calculated based on the traditional characterization of firm technology portfolios using patent (sub-)classes, the correlation with firm performance is much smaller and statistically insignificant. As we argue that a unique and differentiated technology portfolio relates to a competitive advantage and superior firm performance, our findings suggest that patent text more accurately characterizes the position and differentiation of firms in technology space.

Our paper makes an important contribution to the strategy and innovation literature. We introduce a new method to characterize firm technology portfolios, to map a firm's unique spatial position in technology space, and to measure the overall differentiation of a firm's technology portfolio relative to other firms. We demonstrate that our method provides a different characterization of firm technology portfolios compared with the conventional approach based on patent classifications and that our method more strongly correlates with firm profitability and market value. Moreover, we also show that technology differentiation is associated with subsequent differentiation from competitors in the product market and a reduction in outgoing technology spillovers to other firms. Although our analysis is restricted to U.S. public firms, we illustrate that our approach also works for firms with only a few patents in their portfolio. As such, our method could presumably also be used to measure the technological position and differentiation of smaller (nonpublic) firms and start-ups. In addition, we show that the method works for firms specialized in a single-product market industry as well as for diversified firms operating across multiple industries. Whereas the economics and strategy literature has predominantly focused on a firm's competitive position and differentiation in the product market (e.g., Hoberg and Philips 2016, Carlson 2023, Guzman and Li 2023), we introduce and illustrate the importance of a firm's competitive position and differentiation in technology space, particularly in R&D-intensive industries and industries with strong product market rivalry. As a result, our contribution paves the way to explore different causal mechanisms that relate technology differentiation to competitive advantage and firm performance and to study different drivers of technology differentiation. We provide open access to our code and data that might be useful to visually map a firm's spatial position in technology space relative to other firms (Stuart and Podolny 1996), to cluster technologically similar firms (Jaffe 1989), to identify firms with a unique and differentiated technology portfolio (Ahuja and Morris Lampert 2001), to more accurately measure technology spillovers between firms (Bloom et al. 2013), to assess the potential for technology synergies in case of mergers and acquisitions or alliances (Rosenkopf and Almeida 2003), or to study the direction and diversification of firms' technology strategies (Silverman 1999).

2. Methodology

2.1. Data and Sample

To obtain firm technology portfolios, we rely on the DISCERN database that dynamically matches U.S. public firms to U.S. patents for the period 1980–2015 (Arora et al. 2021).[4] We start from a sample of 1,345,945 U.S. patents granted between 1980 and 2015 that are ultimately assigned to any public U.S. firm. In order to construct the technology portfolio of firm i in year t, we collect all patents owned by firm i with a filing year between year $t-5$ and year $t-1$ (Ahuja and Katila 2001, Rothaermel and Deeds 2004, Hirshleifer et al. 2018). For each firm, we obtain data on the firm's position and differentiation in product market space from Hoberg and Philips (2016)[5] and additional financial information from Compustat. Next, we use the processed, cleaned, and stemmed technical keywords extracted from the titles, abstracts, and claims of each U.S. patent from Arts et al. (2021).[6]

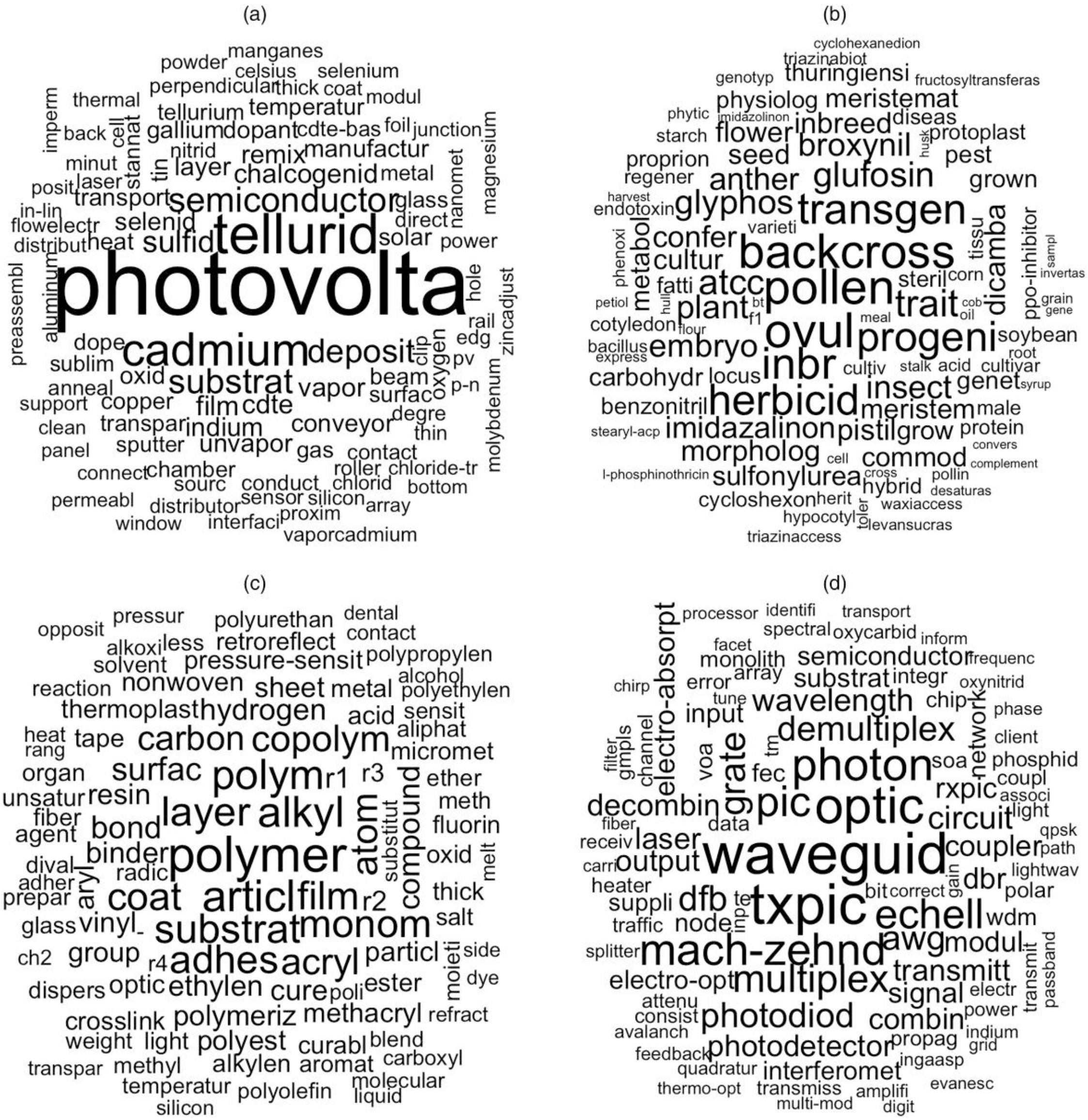
2.2. Characterizing Firm Technology Portfolios and Measuring Technology Similarity

To map each firm's unique spatial position in technology space, we represent the technology portfolio of firm i in year t as a vector of 1,030,335 dimensions, where each dimension corresponds to one stemmed technical keyword from the entire patent vocabulary and each entry captures the share of patents from firm i's patent portfolio in year t that contain the particular keyword. To illustrate this approach, Figure 1 displays word clouds based on the 100 most frequent stemmed technical keywords in the patent portfolio of four different companies. For First Solar, manufacturer of thin-film cadmium telluride photovoltaic panels, the most common stemmed technical keywords include *photovolta*, *cadmium*, and *tellurid*. For Infinera Corporation, manufacturer of wavelength division multiplexing-based packet optical transmission equipment, the most common stemmed technical keywords include *waveguid*, *multiplex*, and *optic*. For Monsanto, the most common stemmed technical keywords include *pollen*, *herbicid*, and *glyphos*, and for 3M, the most common stemmed technical keywords include *adhes*, *polymer*, and *layer*. The average firm-year-level technology portfolio in our sample includes 1,658 unique stemmed keywords,[7] providing a detailed insight in the firm's technology portfolio.[8]

Using these firm-year-level vectors representing the technology portfolio of a firm at a point in time, we compute for each year the technological similarity (*tech similarity*) for every pair of firms by means of cosine similarities (see also Jaffe 1986, 1989). To do so, we use term frequency-inverse document frequency (tf-idf)[9] weights that offset the share of patents containing the keyword in a particular firm-year technology portfolio by the share of all firm technology portfolios from the entire population in a given year that contain the particular keyword.[10] This helps to adjust both for the fact that some keywords are more representative of a firm's technology portfolio in a given year (e.g., *photovolta* for First Solar or *herbicid* for Monsanto) and for the fact that some keywords appear frequently across the technology portfolio of many firms in a given year and are therefore less discriminating across firms (e.g., *electr* or *drug*). As a result, new or emerging technologies as captured by new (or recent) keywords introduced for the first time in history and that occur less frequently across the technology portfolios of all firms receive a higher weight (Arts et al. 2021). We calculate for each year the *tech similarity* for every pair of firms. The final data set covers the years 1980–2015 and includes 4,832 firms, 57,772 firm-year observations, and 98,279,118 pairwise *tech similarities* for all firm-year observations. As expected, the average *tech similarity* between firms is very low (mean = 0.057, median = 0.034, standard deviation = 0.069), and only a small share of all firm pairs displays a high degree of *tech similarity*. Our new text-based *tech similarity* between firms is only moderately correlated with the traditional similarity measures based on patent classifications (e.g., Jaffe 1989).[11] As an illustration, Table A.2 in the online appendix ranks, for a selected number of focal firms, the top 10 most similar firms according to different measures. There is only a moderate overlap between the firms identified by *tech similarity* and those identified by the traditional metrics. As such, our new approach based on patent text and tf-idf provides a different characterization of firm technology portfolios and a firm's spatial position in technology space compared with the traditional approach based on patent classifications.[12] Interestingly, the correlation between the similarity between firms in technology space and the similarity between firms in product market space, based on the product descriptions from annual 10-K reports, is only 0.25 (Hoberg and Philips 2016).[13] This relatively low correlation illustrates that companies competing with similar products in the same industry often rely on different types of technology and/or that the same technologies are used by firms from different industries (Jaffe 1986, 1989). The companies with the most similar technology portfolios in history include IBM and Digital Equipment in 1994 (*tech similarity* = 0.904), Baker Hughes and Schlumberger (both providing oil field services) in 2004 (*tech similarity* = 0.906), AT&T and Sprint in 2006 (*tech similarity* = 0.906), Alphabet (Google) and Altaba (Yahoo!) in 2009 (*tech similarity* = 0.931), and Texas Instruments and Freescale Semiconductor in 2012 (*tech similarity* = 0.923).

Despite its simplicity and benefits, tf-idf has a number of important limitations. It fails to account for synonyms (different words with same meaning), polysemy (same word with different meanings), and the order and context of words in patent documents. To both test the robustness of our approach based on tf-idf and offer an alternative representation of a firm's technology portfolio that does account for synonyms, polysemy, and the

Figure 1. Characterizing Firm Technology Portfolios Using Word Clouds



Notes. This figure illustrates a simplified version of our characterization of firm technology portfolios using word clouds for four companies, namely First Solar, Monsanto, 3M, and Infinera Corporation. For each firm, we plot its top 100 most frequently used stemmed technical keywords identified from its patents granted between 1980 and 2015; word size is proportional to tf-idf weights. (a) First Solar. (b) Monsanto. (c) 3M. (d) Infinera Corporation.

order and context of words in patents, we train a doc2vec model on the entire corpus of U.S. patents,[14] create a document-embedding vector for every patent, and average the patent-level vectors for all patents in a firm's technology portfolio in a given year to map a firm's spatial position in technology space in a given year (Le and Mikolov 2014, Guzman and Li 2023). Next, we use this alternative representation of a firm's technology portfolio to calculate *tech similarity* (*doc2vec*) for every pair of firms and each year by means of cosine similarities (n = 98,279,118). The correlation between *tech similarity* and *tech similarity* (*doc2vec*) is 0.70. In the remainder of the paper, we will use the most straightforward approach based on tf-idf. As illustrated later in the paper, our findings hold if we use doc2vec instead of tf-idf. We did not use Bidirectional Encoder Representations from Transformers (BERT) to characterize the technology portfolios of firms because BERT is better suited for short text documents of a maximum of 512 tokens (Carlson 2023).

Our open access code and data can be used to measure and visualize each firm's unique spatial position in technology space relative to all—or to a selected number of—other firms, and to cluster firms based on the similarity of their technology portfolio in line with the text-based product industry clustering of Hoberg and Philips (2016). Scholars can use our input data and code together with their preferred clustering and visualization methods, such as *k*-means or hierarchical clustering. As an illustration, Figure 2 shows a network graph restricted to all firms in the machinery industry in 2005. Each node represents one firm, the size of the node is proportional to the size of the firm's patent portfolio in 2005 (based on patents from 2000 to 2004), two nodes are connected by an edge in the case that *tech similarity* between the firms is above 0.5, and the thickness of the edge is proportional to *tech similarity* between the firms (a thicker edge means higher *tech similarity*). The colors represent six clusters of technologically similar firms based on hierarchical clustering using the Ward method: semiconductor equipment firms (e.g., Applied Materials); engine technology firms (e.g., Caterpillar); fluid and air filter technology firms (e.g., Donaldson Company); mining, oil, and gas drilling technology firms (e.g., Baker Hughes); photonics and laser technology firms (e.g., Veeco Instruments); and power tools and component technology firms (e.g., Black & Decker).[15]

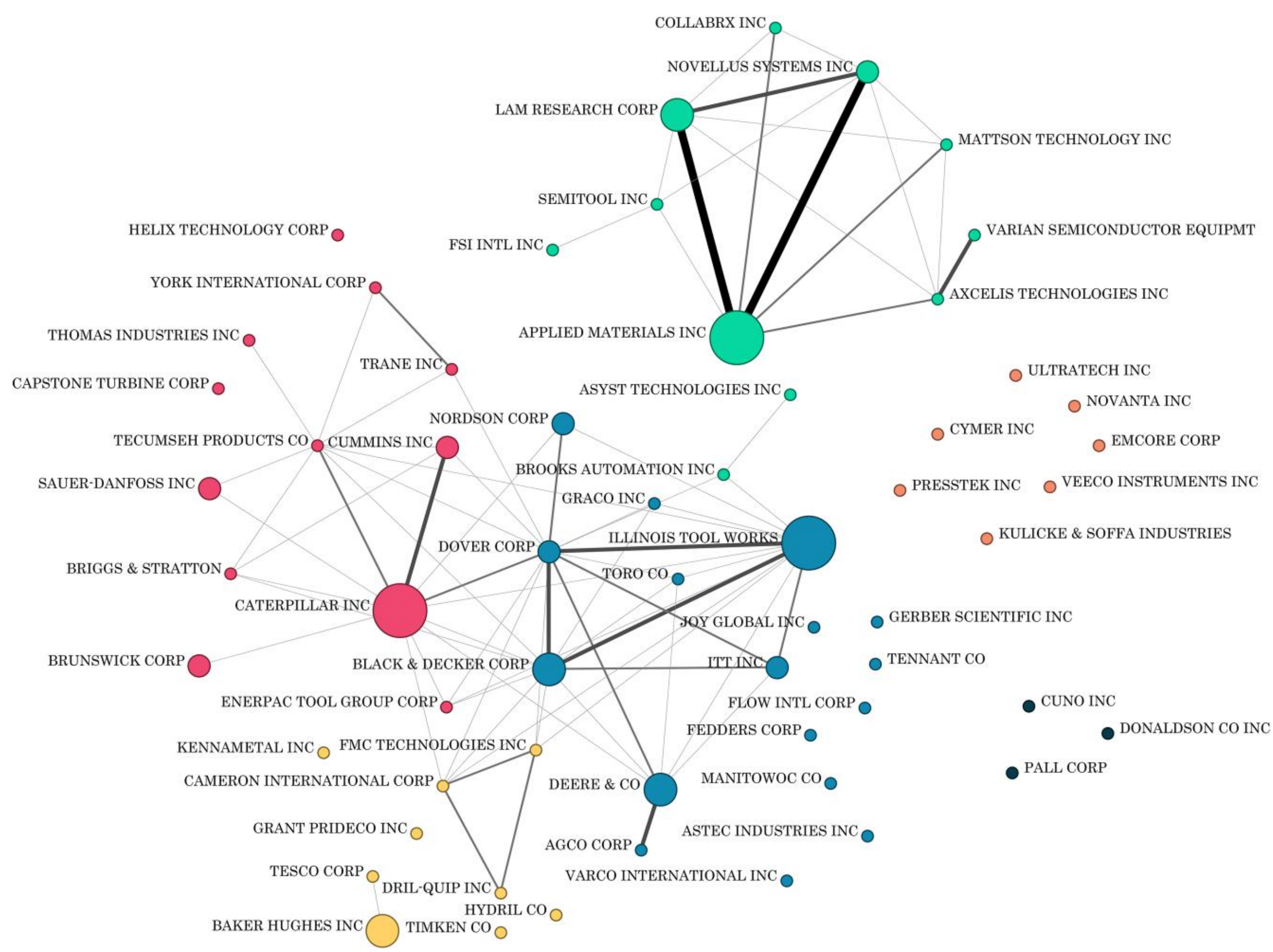
2.3. Technology Differentiation

To measure the technology differentiation of firm i in year t, we calculate

$$\textit{tech differentiation}_{it} = 1 - \frac{1}{n-1}\sum_{j=1,\ j\neq i}^{n} \textit{tech similarity}_{ijt},$$

with n equal to firms active in year t and $\textit{tech similarity}_{ijt}$ equal to the *tech similarity* between firm i and firm j in year t. In calculating $\textit{tech differentiation}_{it}$, we only include the 10% most similar firms in technology space to firm i in year t as we expect technology competition to play out

Figure 2. (Color online) Technology Position of Firms in the Machinery Industry in 2005



Notes. This network graph displays all firms from the machinery industry (as defined by Fama and French 1997) in 2005 with at least 20 granted patents filed between 2000 and 2004. Nodes represent firms, and node size is proportional to the number of granted patents filed by the firm between 2000 and 2004. Two nodes are connected by an edge if *tech similarity* is larger than 0.5, and edge thickness is proportional to *tech similarity* between firms (thicker edges mean more similar firms). Node colors indicate clusters of similar firms based on hierarchical clustering using the Ward method. Six clusters are identified, but this number is set and not determined by the algorithm. The green cluster largely corresponds to semiconductor equipment technology firms (e.g., Applied Materials); the red cluster corresponds to engine technology firms (e.g., Caterpillar); the black cluster corresponds to fluid and air filter technology firms (e.g., Donaldson Company); the yellow cluster corresponds to mining, oil, and gas drilling technology firms (e.g., Baker Hughes); the orange cluster corresponds to photonics and laser technology firms (e.g., Veeco Instruments); and the blue cluster corresponds to power tools and component technology firms (e.g., Black & Decker).

between the most technologically similar firms. Nevertheless, all findings and robustness checks in the paper remain consistent if we use a broader or more narrow sample of firms to calculate *tech differentiation*$_{it}$ using either all other firms or only the 5% most similar firms to firm i in year t. In fact, because the distribution of *tech similarity*$_{ijt}$ has a very low mean and a long right tail, *tech differentiation*$_{it}$ is predominantly driven by the firms most similar to firm i in year t.[16] To increase technology differentiation, firms can internally develop or externally acquire new or emerging technologies, which by definition, are different from the technology from virtually all other firms, or internally develop or externally acquire less novel technologies that are nonetheless different from their closest neighboring firms in technology space. The correlation between *tech differentiation* of firm i in year t and the number of new keywords appearing for the first time in history in a patent and pioneered by firm i in the same period is 0.14, illustrating how firms pioneering new technologies can increase their technology differentiation.[17] For instance, in 2000, McAfee's technology portfolio became more differentiated by being the first to introduce the keyword *malware* in a patent. Overture Services filed a patent in 2003 for "a method and system for optimum placement of advertisements on a webpage," the first patent in history introducing *cost-per-click* (i.e., the internet advertising technology used by Google and others). The company was acquired by Yahoo! in the same year. The latter example illustrates how firms can increase their technology differentiation not only from internal R&D but also, by acquiring external technology through among others, mergers and acquisitions (Arora and Gambardella 1990, Arora et al. 2001). Therefore, technology differentiation might play a key role in M&A transactions that has been overlooked by prior empirical studies on the role of innovation for M&As (Arts et al. 2022).

Tech differentiation can vary between industries, between firms within the same industry, and within a firm over time. As displayed in the variance decomposition in Table A.3 in the online appendix, differences between industries account for approximately 22% of the total variance in *tech differentiation*, differences between firms within the same industry account for 61% of the variance, and variation in *tech differentiation* within the same firm over time accounts for the remaining 17%. As expected and as illustrated in the correlation matrix in Table A.4 in the online appendix, younger, smaller, more R&D-intensive, and more technology-specialized firms display higher levels of technology differentiation relative to older, larger, less R&D-intensive, and less technology-specialized firms. Firms with the highest degree of technology differentiation in history across all industries include, for example, Pioneer Hi-Bred International in 1997 or Monsanto in 2011. Companies with a persistent high level of technology differentiation in their respective industries over time include, for example, Monsanto (agricultural production and crops), Tesla (motor vehicles and passenger car bodies), Olin (inorganic chemicals), or First Solar (semiconductors and related devices).[18]

To compare the use of patent text versus patent classification to measure technology differentiation, we calculate *tech differentiation (class)* and *tech differentiation (subclass)* in the same way except for using patent classes or subclasses instead of patent text.[19] Interestingly, our *tech differentiation* measure based on text only weakly correlates with *tech differentiation (class)* (correlation = 0.02) and *tech differentiation (subclass)* (correlation = 0.14). Tables A.5 and A.6 in the online appendix rank the top 10 firms with the most differentiated technology portfolio in history across all industries and by selected industries and years, respectively, and again, they indicate that patent text offers a very different insight in a firm's position and differentiation in technology space compared with patent classifications.

3. Technology Differentiation and Firm Performance

To validate the contribution of our text-based measure of technology differentiation to the strategy and innovation literature, we examine the relation between our measure of technology differentiation and firm performance. Next, we study whether—as expected—technology differentiation is more important for firm performance in R&D-intensive industries and in industries with stronger product market rivalry. Finally, we empirically examine the two mechanisms advanced by theory for why technology differentiation might increase firm performance. Specifically, we test whether technology differentiation is associated with subsequent differentiation from competitors in the product market and a reduction in outgoing technology spillovers to other firms.

3.1. Data and Sample

To study how technology differentiation relates to firm performance, we use the entire firm-year-level panel of U.S. public firms matched to patents and collect *Tobin's Q*[20] as a proxy for firm performance from Compustat. In line with prior work, we use *Tobin's Q* as a measure of firm performance because the market value of a firm is taking the effect of technology differentiation on future sales and profits into consideration (e.g., Hall et al. 2005, Simeth and Cincera 2016, Bellstam et al. 2021). As illustrated in the online appendix, our main findings continue to hold if we use return on assets (*ROA*) as an alternative measure of firm performance. In line with prior work (e.g., Hirshleifer et al. 2018, Bellstam et al. 2021), we use *total assets*, *leverage*, *cash*, *asset tangibility*, and *firm age* as control variables. In addition, we control for *R&D intensity* (R&D investments/total assets) and *citation-weighted patents* (number of patents in the portfolio of firm i in year t (i.e., filed between $t-5$ and $t-1$) weighted by the number of forward cites received by these patents). To control for the technology specialization of firms, we include both *tech class count* (number of distinct main patent classes linked to the patents in the portfolio of firm i in year t) and *tech class concentration* (the degree to which the patent portfolio of firm i in year t is concentrated in a small number of technology classes).[21] Finally, given the important relation between product market competition and firm performance and particularly between product market competition and firm-level innovation strategy (Aghion et al. 2005, Igami and Uetake 2020), we additionally control for the yearly amount of *prod market competition* faced by the focal firm in year t using a Hirschman–Herfindahl index based on the sales of all firms within the same industry in year t reported in Compustat. In line with prior studies, we winsorize all variables from Compustat (except for *firm age*) at the 1% and 99% levels (e.g., Custódio et al. 2019). We end up with an unbalanced panel of 4,754 firms and 52,418 firm-year-level observations for the years 1980–2015. A firm is on average observed over approximately 11 years.[22] All variables are defined in Table A.1 in the online appendix. Table 1 shows descriptive statistics, and Table A.4 in the online appendix displays the correlation matrix.

3.2. Results

We study how technology differentiation relates to firm performance using the following firm-year-level panel model:

$$Y_{it} = \alpha_i + \gamma_j + \delta_t + \beta_1 \textit{ tech differentiation}_{it} + \beta_2 X_{it-1} + \varepsilon_{it}. \quad (1)$$

The performance of firm i in year t (Y_{it}) is measured by *Tobin's Q*; α_i, γ_j, and δ_t capture firm, industry,[23] and year fixed effects; and X_{it-1} includes all control variables lagged by one year (i.e., *total assets*, *leverage*, *cash*, *asset tangibility*, *firm age*, *R&D intensity*, *citation-weighted patents*, *tech class count*, *tech class concentration*, and *prod market competition*). Notice that *tech differentiation*$_{it}$ is measured based on firm i's patents from years $t-5$ to $t-1$.[24] Firm and industry fixed effects control for time-invariant unobserved heterogeneity across firms and industries, and year fixed effects control for unobserved heterogeneity across years. In Table A.8 in the online appendix, we illustrate the robustness of our main findings if we additionally control for unobserved heterogeneity at the industry-year and technology(-year) levels.[25]

As shown in Table 2, *tech differentiation* has a strong positive relation with firm performance significant at the 1% level both in the cross-section (column (1)) and with firm fixed effects (column (2)). Using our preferred model with firm fixed effects, a one-standard deviation increase in *tech differentiation* is related to an increase in *Tobin's Q* of 10.4%. Using *ROA* as an alternative measure for firm performance, we find that a one-standard deviation increase in *tech differentiation* is related to a 2.0% increase in *ROA* (Table A.9 in the online appendix). Using doc2vec instead of tf-idf to measure each firm's position and differentiation in technology space, a one-standard deviation increase in *tech differentiation* (*doc2vec*) is related to a 6.9% increase in *Tobin's Q* (Table A.10 in the online appendix). The effect of *tech differentiation* on firm performance is also significant in comparison with the effects of the traditional firm innovation measures that capture a firm's rate of invention. A one-standard deviation increase in *R&D intensity* and *citation-weighted patents* affects *Tobin's Q* by 4.3% and 18.3%, respectively. Thus, the differentiation of a technology portfolio is arguably an important driver of firm performance next to the size of the technology portfolio as measured by the

Table 1. Descriptive Statistics

	Variables	Mean	Median	Standard deviation	Min	Max
(1)	*Tobin's Q*	2.447	1.485	3.130	0.115	22.612
(2)	*ROA*	−4.640	9.605	45.940	−292.807	37.272
(3)	*Tech differentiation*	0.817	0.844	0.102	0.415	0.989
(4)	*Tech differentiation (doc2vec)*	0.481	0.496	0.135	0.136	0.766
(5)	*Tech differentiation (class)*	0.696	0.729	0.167	0.182	1.000
(6)	*Tech differentiation (subclass)*	0.954	0.970	0.047	0.267	1.000
(7)	*R&D intensity*	12.634	5.718	20.706	0.000	133.681
(8)	*Citation-weighted patents*	678.804	47.000	3,920.849	1.000	170,542.000
(9)	*Tech class concentration*	0.457	0.368	0.325	0.013	1.000
(10)	*Tech class count*	21.819	8.000	38.199	1.000	373.000
(11)	*Prod market competition*	0.143	0.095	0.138	0.015	1.000
(12)	*Total assets*	1,855.184	129.256	5,866.021	0.618	41,575.000
(13)	*Leverage*	20.763	14.534	26.693	0.000	178.273
(14)	*Cash*	25.579	15.299	26.275	0.077	95.026
(15)	*Asset tangibility*	21.548	18.091	16.291	0.208	72.833
(16)	*Firm age*	11.817	9.000	8.637	0.000	39.000

Notes. This table reports the descriptive statistics of the sample used to examine the relationship between *tech differentiation* and firm performance from 1980 to 2015, and it includes 52,418 firm-year observations and 4,754 firms. *ROA*, *R&D intensity*, *Leverage*, *Cash*, and *Asset tangibility* are measured as percentages. We set missing values for *R&D intensity*, *Leverage*, *Cash*, and *Asset tangibility* to zero. All financial measures from Compustat are winsorized at levels of 1% and 99%. Definitions of variables can be found from Table A.1 in the online appendix.

Table 2. Technology Differentiation and Firm Performance

	(1)	(2)	(3)	(4)
Tech differentiation	0.579***	0.964***		
	(0.133)	(0.131)		
Tech differentiation (class)			0.134	
			(0.085)	
Tech differentiation (subclass)				0.191
				(0.139)
R&D intensity	0.734***	0.204***	0.196***	0.194***
	(0.047)	(0.045)	(0.045)	(0.045)
Citation-weighted patents	0.109***	0.084***	0.065***	0.065***
	(0.008)	(0.009)	(0.008)	(0.008)
Firm fixed effects	No	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes
Other control variables	Yes	Yes	Yes	Yes
Number of observations	51,316	51,316	51,316	51,316
Number of firms	4,741	4,741	4,741	4,741
Within R^2		0.164	0.162	0.161
Between R^2		0.179	0.176	0.177
Overall R^2	0.242	0.132	0.128	0.129
Marginal effects (%)				
Tech differentiation	6.11	10.37		
Tech differentiation (class)			2.26	
Tech differentiation (subclass)				0.91
R&D intensity	16.25	4.27	4.10	4.06
Citation-weighted patents	24.26	18.29	13.79	13.94

Notes. The table reports coefficient estimates from a linear firm fixed effects regression (except column (1), which is standard ordinary least squares). The sample is an unbalanced panel with firm fiscal years ranging from 1980 to 2015. *Tobin's Q* and *Citation-weighted patents* are log transformed. Additional control variables include *Total assets* (log), *Firm age* (log), *Leverage*, *Cash*, *Asset tangibility*, *Tech class concentration*, *Tech class count* (log), and *Prod market competition*. We set missing values for *R&D intensity*, *Leverage*, *Cash*, and *Asset tangibility* to zero. Control variables are lagged by one year. All financial measures from Compustat are winsorized at levels of 1% and 99%. Industry fixed effects are based on three-digit Standard Industrial Classification (SIC). Definitions of variables are provided in Table A.1 in the online appendix. Robust standard errors (clustered at the firm level) are reported in parentheses. Marginal effects indicate the change of dependent variable related to a one-standard deviation increase of the corresponding explanatory variable.

***Significance at the 1% level.

number of *citation-weighted patents*. Notice that *citation-weighted patents* relies on future information in regard to the impact and likely commercial value of individual patents, which only becomes available many years after a patent is granted, whereas *tech differentiation* exclusively relies on the technical content of patents at the time they are granted (Hall et al. 2005). In line with prior work in this stream of research, we have no exogenous variation, and accordingly, our results should be interpreted as new descriptive facts rather than causal estimates (e.g., Hall et al. 2005, Simeth and Cincera 2016, Bellstam et al. 2021).

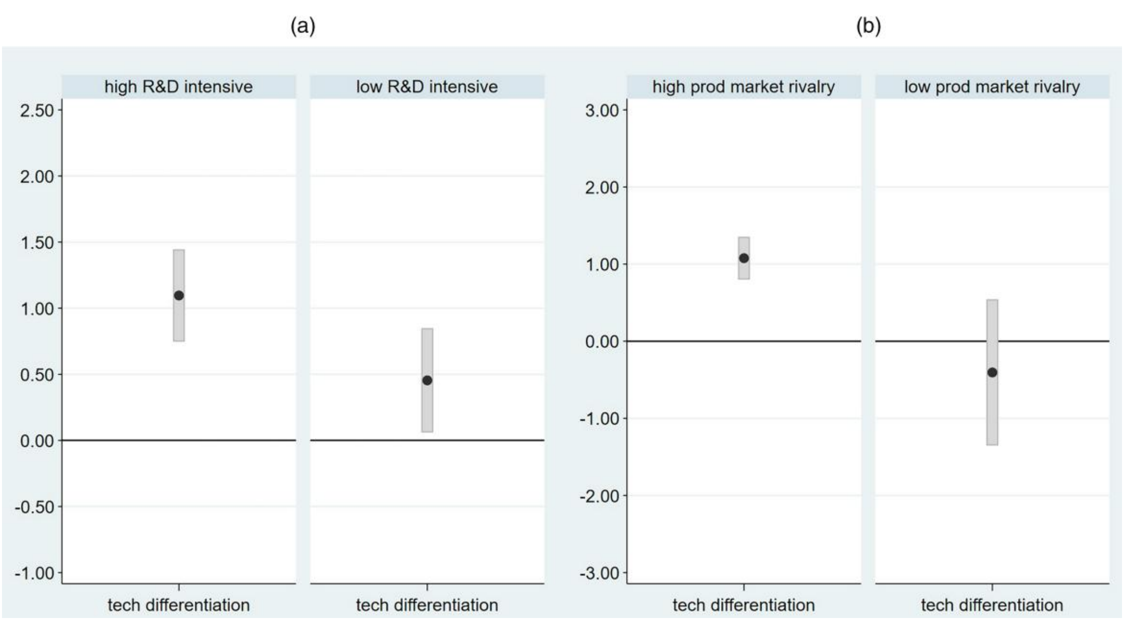
As illustrated in the online appendix, our *tech differentiation* measure also works for firms with only a few patents in their portfolio as well as for firms with a large patent portfolio (Table A.11 in the online appendix). Because a single patent document has on average 61 unique stemmed technical keywords, a few patents can already provide a good insight into a firm's unique position and differentiation in technology space. As such, our method could arguably also be used for smaller (nonpublic) firms and start-ups: for instance, by replicating our method and code on the Orbis Intellectual Property database, which links a larger sample of both public and private firms to patent data. The method also works for firms specialized in a single-product market industry as well as for diversified firms operating in multiple industries (Table A.12 in the online appendix). In addition, our main findings also remain robust if we calculate technology differentiation exclusively based on a comparison with all other firms or only the top 5% or single closest firm(s) in technology space (instead of the top 10%). We also find consistent results if we calculate technology differentiation based on the top 10%, top 5%, or single closest firm(s) in product market space (i.e., by measuring how differentiated the technology portfolio of the focal firm is from its closest product market rivals) (Table A.13 in the online appendix). Moreover, given that our measure of technology differentiation is based on patent text and that many patents are written by patent lawyers who may strategically draft patents to maximize the scope of invention or to strategically obscure the invention (Arinas 2012, Sampat 2018, Kong et al. 2023), we include patent lawyer or law firm fixed effects to test the robustness of our method.[26] Table A.14 in the online appendix shows that *tech differentiation* continues to have a strong positive relation with *Tobin's Q* after controlling for individual lawyer or law firm fixed effects. Finally, if we calculate technology differentiation based on patent classes or subclasses rather than patent text, we find that neither *tech differentiation (class)* nor *tech differentiation (subclass)* have a statistically significant relation with firm performance. Their marginal effects on *Tobin's Q* are very small and not significantly different from zero (Table 2, columns (3) and (4)). *Tech differentiation (class)* and *tech differentiation (subclass)* are statistically insignificant even if we do not control for firm fixed effects (not shown in Table 2). Moreover, both *tech differentiation (class)* and *tech differentiation (subclass)* have a statistically insignificant effect on *ROA* (Table A.9 in the online appendix).

3.3. Industry R&D Intensity and Rivalry

To better understand the relation between technology differentiation and superior firm performance, we split the sample by the R&D intensity and product market rivalry of an industry. In R&D-intensive industries, firms heavily invest in R&D because pushing the technology frontier and creating a differentiated technology portfolio are crucial for a firm's competitive advantage and financial performance. Therefore, we expect a stronger relation between technology differentiation and firm performance in industries with high R&D intensity, such as medical equipment or turbine and engines, in comparison with industries with a lower R&D intensity, such as the food or furniture manufacturing industry. To test this, we collect the annual R&D intensity of all U.S. public firms from Compustat, calculate *industry R&D intensity* by taking the average *R&D intensity* of all firm-year observations from the same industry, and split the sample by the mean of *industry R&D intensity*. As illustrated in panel (a) of Figure 3,[27] *tech differentiation* has a significantly stronger relation with *Tobin's Q* for firms in high R&D-intensive industries compared with firms in industries with a lower R&D intensity ($Z = 2.392$, $p = 0.017$). A one-standard deviation increase in *tech differentiation* corresponds with an increase in *Tobin's Q* of 12.0% (significant at $p = 0.000$) for firms in high R&D-intensive industries versus 4.5% (significant at $p = 0.024$) for firms in low R&D-intensive industries. We find similar results if we split the sample of industries by the median instead of the mean of *industry R&D intensity* (Table A.16 in the online appendix).

Likewise, in industries with strong product market rivalry, such as the computer hardware or the fabricated metal products manufacturing industry, firms have a stronger incentive to push the technology frontier and create a differentiated technology portfolio in order to escape competition compared with firms in industries with less product market rivalry, such as the beverage, tobacco, or musical instruments industry (Aghion et al. 2005, Hoberg and Philips 2016). In case many firms sell similar products to the same customers, firms arguably

Figure 3. (Color online) Technology Differentiation and Firm Performance by Industry Type



Notes. Panels (a) and (b) illustrate estimated coefficients and the corresponding 95% confidence intervals of *tech differentiation* on *Tobin's Q* for samples split by the mean of *industry R&D intensity* and *industry prod market competition*, respectively. Regression results can be found in columns (1)–(4) in Table A.15 in the online appendix. (a) Industries with high vs. low R&D intensity. (b) Industries with high vs. low product market rivalry.

face a greater risk of losing business to firms with a differentiated technology portfolio. Therefore, we expect a stronger relation between technology differentiation and firm performance in industries with stronger product market rivalry relative to industries with weaker product market rivalry. To test this, we calculate for every industry in Compustat the average *prod market competition* (Hirschman–Herfindahl concentration index) across years, and we split the sample of industries based on the average degree of product market competition. Panel (b) of Figure 3 shows that *tech differentiation* has a significantly stronger relation with *Tobin's Q* for firms in industries with high product market rivalry relative to firms in industries with low product market rivalry ($Z = 2.959$, $p = 0.003$).[28] A one-standard deviation increase in *tech differentiation* relates to an increase in *Tobin's Q* of 11.7% (significant at $p = 0.000$) for firms in industries with high product market rivalry versus −3.2% (insignificant at $p = 0.401$) for firms in industries with lower product market rivalry. We find similar results if we split the sample of industries by the median instead of the mean of product market competition (Table A.16 in the online appendix). As an additional robustness check, we use two alternative metrics for the degree of product market rivalry in an industry. First, we use data from Hoberg and Philips (2016) and calculate a proxy for the product market rivalry of an industry using the yearly similarities in the 10-K business descriptions of U.S. public firms from the same industry.[29] Second, we collect the Hirschman–Herfindahl index for the concentration of each industry from the 2017 U.S. Census, which includes both public and private firms. As illustrated in Tables A.15 and A.16 in the online appendix, we continue to find that *tech differentiation* has a significantly stronger relation with firm performance for firms in industries with stronger product market rivalry.[30]

3.4. Product Market Rivalry and Technology Spillovers as Potential Mechanisms

The theory advances two potential mechanisms that could drive the effect of technology differentiation on firm performance. First, unique and difficult to substitute or imitate resources, such as a unique and proprietary technology portfolio, might create market power and reduce rivalry in the product market, and thereby, they might increase firm performance (Wernerfelt 1984, Peteraf 1993, Mowery et al. 1998). Firms with a proprietary and differentiated technology portfolio can arguably develop unique products or processes that appeal to customers, create market power, and make it more difficult for rivals to enter the same product market space (Sutton 1991, Hoberg and Philips 2016). Following Hoberg and Philips (2016), we calculate product market overlap as the sum of product market similarities between the focal firm and its top 1%, top 5%, or top 10% closest rivals in the product market. As illustrated in the firm fixed effects models in Table A.17 in the online appendix, a standard deviation increase in technology differentiation from product market rivals is associated with a significant 2.3%–4.7% decrease in the product market overlap between the focal firm and these product market rivals.

Second, firms might differentiate their technology portfolio to minimize outgoing technology spillovers to other firms and potential competitors and as a result, reduce competition in the product market and increase firm performance (Kamien and Zang 2000, Cassiman and Veugelers 2002, Aghion et al. 2005, Gil Moltó et al. 2005, Lin and Zhou 2013). As shown in our preferred specification in column (4) of Table A.18 in the online appendix (firm fixed effects Poisson quasimaximum likelihood), a standard deviation increase in *tech differentiation* is associated with a 16.7% reduction in outgoing technology spillovers to other firms as measured based on patent citations (Jaffe et al. 1993). Although citations are a noisy measure of spillovers (Jaffe et al. 2000), these results are consistent with formal theory models suggesting that firms tend to differentiate their technology from other firms to minimize outgoing spillovers.

4. Discussion and Conclusion

Whereas the rate of invention by firms and the effect on firm performance have been central themes in economics and strategy (Hall et al. 2005, Bellstam et al. 2021), the position and differentiation of invention by firms have received far less attention in the literature. To contribute to this stream of research, we develop a new method to characterize a firm's technology portfolio based on the semantic content of patents, to map each firm's unique spatial position relative to every other firm in technology space, and to measure the overall differentiation of a firm's technology portfolio. We show that our method provides a different characterization of firm technology portfolios compared with the traditional approach based on patent classifications and more strongly correlates with firm profitability and market value. Moreover, although the economics and strategy literature has mainly focused on a firm's competitive position and differentiation in the product market (Hoberg and Philips 2016, Carlson 2023, Guzman and Li 2023), we introduce and empirically demonstrate the importance of a firm's competitive position and differentiation in technology space. Using a panel of all U.S. public firms matched to patents for the years 1980–2015, our findings suggest that a firm's competitive advantage and superior performance rely on a unique and differentiated technology portfolio, particularly in R&D-intensive industries and in industries with strong product market rivalry. We also show that technology differentiation is associated with subsequent differentiation from competitors in the product market and a reduction in outgoing technology spillovers to other firms. Although we do not fully causally identify the mechanisms relating technology differentiation to firm performance, we hope that our work unlocks different avenues for future research in this area. Finally, we provide open access to all code and data to characterize the technology portfolio of firms and to measure the technological similarity and differentiation of U.S. public firms between 1980 and 2015.

Acknowledgments

Comments from Juan Alcacer, Paul Anckaert, Ashish Arora, Andrea Fosfuri, Tim Simcoe, Giovanni Valentini, David Wehrheim, and David Yoffie as well as seminar participants at Boston University, Cornell University, Harvard Business School, the National Bureau of Economic Research, National Tsing Hua University, the UK Intellectual Property Office, Utrecht University, the University of Pisa, the 2021 Annual Meeting of the Academy of Management, the 2021 Strategic Management Society conference, and the 2021 Druid conference improved the paper substantially. Research assistance from Nicola Melluso with doc2vec and Lena Veretennik with network graphs is acknowledged.

Endnotes

[1] Data are available from https://zenodo.org/record/5172146, and code is available from https://github.com/JiananHou429/measuring-the-position-of-firms-in-technology-space.

[2] See 35 U.S. Code §112, which is available from https://www.law.cornell.edu/uscode/text/35/112.

[3] Using expert assessments (Arts et al. 2018) and patent examiner assignments (Righi and Simcoe 2019), prior work illustrates that there remains substantial heterogeneity in the technological content and similarity of patents assigned to the same (sub-)classes and that patent text more accurately measures the content and similarity of patents.

[4] The DISCERN database is available from https://zenodo.org/record/3709084. DISCERN dynamically matches each patent assignee to its ultimate owner in different time periods by combining multiple sources, such as Orbis, Securities Data Company Platinum, Center for Research in Security Prices monthly stock, and the National Bureau of Economic Research patent database, in order to account for firm name changes, subsidiaries, and ownership changes because of mergers and acquisitions.

[5] Available from http://hobergphillips.tuck.dartmouth.edu/tnic_poweruser.htm.

[6] For each patent, they concatenate title, abstract, and claims; lowercase text; and tokenize it to words using the following regular expression: [a-z0-9][a-z0-9-]*[a-z0-9]+|[a-z0-9]. They consider a word as a sequence of letters and numbers that could be separated by hyphens ("-"). Next, they remove words composed only of numbers, one-character words, stop words from the Natural Language Toolkit in the Python library, and words appearing in only one patent. In addition to natural stop words, they remove a manually compiled list of 32,255 very common nontechnical keywords. Finally, they apply stemming to each word using the SnowBall method. The entire cleaned vocabulary contains 1,362,971 unique stemmed keywords, and the average and median numbers of unique stemmed keywords per patent granted since 1976 are 61 and 56, respectively.

[7] For instance, we retrieved 1,671 stemmed keywords from 135 First Solar patents, 22,743 from 6,895 Monsanto patents, 41,078 from 15,237 3M patents, and 2,247 from 295 Infinera patents.

[8] Although our approach to characterize the technology portfolio of firms arguably captures the combination of all technologies created by the firm, it does not capture whether a firm combines these technologies for the creation of a single invention or patent (Yayavaram and Ahuja 2008).

[9] The tf-idf adjustment is conducted as follows. First, we construct a vector for each firm-year where the value of each dimension captures the term frequency (tf), namely the share of patents in the firm portfolio using the given word. Second, for each word, we calculate its inverse document frequency (idf), namely the total number of firms in the sample divided by the number of firms using this given word. Because of the high skewness of idf, we take the logarithm with base 10. Finally, the adjusted value equals the product of tf and idf. The tf-idf weights are calculated for each firm-year observation. To calculate tf-idf, we use shares rather than counts because of the large differences in the length of patent documents (title, abstract, claims) and in the number of patents linked to a firm-year-level technology portfolio.

[10] Alternatively, one might group all patents in the portfolio of firm i in year t together and treat them as one text document representing the technology portfolio of firm i in year t. We did not follow this alternative approach because it ignores the patent level as signaling the importance of a given technology for a firm, and it would treat one patent mentioning a keyword n times the same as n patents mentioning a keyword only once. Moreover, patent documents (title, abstract, claims) vary greatly in length. Some patent documents are very short, whereas others cover many pages. Grouping patent documents together would cause longer patent documents to have a bigger influence on the representation of a firm's technology portfolio compared with shorter patent documents. Yet, the downside of the approach used in the paper is that it does not account for the frequency of a keyword within a single patent document.

[11] In order to compare our new text-based *tech similarity* measure with traditional metrics, we calculated *tech similarity (class)* and *tech similarity (subclass)* in the exact same way except for using all of a patent's United States Patent Classification classes or subclasses instead of keywords to characterize firm patent portfolios. Table A.1 in the online appendix provides a detailed overview of how each measure is calculated. The new text-based *tech similarity* measure has only a moderate correlation with *tech similarity (class)* (correlation = 0.39) and *tech similarity (subclass)* (correlation = 0.32).

[12] Technological language might evolve over time. However, we are comparing firms at the same point in time and will control for time fixed effects in the analyses.

[13] We downloaded product similarity scores between 1988 and 2014 from the Text-based Network Industry Classifications database (Hoberg and Philips 2016). The similarity is only available for firm-years in which 10-K filings are available and are covered by Compustat. Financial firms (SIC 6000–6999) and firm-years with nonpositive sales or with assets less than 1 million are excluded.

[14] To generate word embeddings, we use the five preceding and five succeeding words along with 700 dimensions to represent the patent-level vector.

[15] As another example, Figure A.1 in the online appendix shows the network of the top 100 firms with the largest patent portfolio in 2000 across all industries.

[16] Although there is no theoretical motivation for restricting attention to exactly the 10% most technologically similar firms, given the skewedness of the distribution of technological similarity, this sample definitely includes all firms that are reasonably close in technology space. The correlations between *tech differentiation* calculated based on a comparison with all other firms and *tech differentiation* calculated exclusively based on a comparison with the top 10% or top 5% closest firm(s) are 0.96 and 0.94, respectively. Moreover, restricting the measure to the top 10% most similar firms in technology space guarantees a reasonable overlap with product market rivals from the same industry.

[17] The correlation is significant at the 1% level. *Tech differentiation* also positively correlates with the share of patents from the firm's patent portfolio without any backward prior art citations (correlation = 0.06, significant at the 1% level), a measure that has been used in prior studies to identify pioneering technologies (Ahuja and Morris Lampert 2001).

[18] Industries are defined based on SIC. The selection is based on *tech differentiation* and restricted to firms with at least 100 patents in their technology portfolio.

[19] Table A.1 in the online appendix provides a detailed description on the calculation of each measure.

[20] All findings remain robust if we use an alternative calculation of Tobin's Q, which better accounts for intangible assets, amongst others by accounting for all prior R&D investments of the firm (see Peters and Taylor 2017).

[21] As expected, technology-specialized firms have a higher degree of *tech differentiation* relative to technology-diversified firms whose technology portfolio by construction has a certain overlap with a larger number of other firms. Nevertheless, the variance inflation factor of *tech differentiation* is 3.5, indicating that multicollinearity is unlikely to significantly bias our estimates.

[22] Our results are robust for firm survivor bias (i.e., for the subset of firms that remained active until the last year in our sample; that is, 2015). Also, all findings remain robust for the subset of firms that are observed over at least 10 years.

[23] Industry fixed effects are generated based on three-digit SIC codes. We also include industry fixed effects in the models with firm fixed effects because a small number of firms changed SIC codes over time. Table A.7 in the online appendix shows that our findings hold if we generate industry fixed effects based on the Fixed Industry Classification developed by Hoberg and Philips (2016), which classifies firms into industries at various levels of granularity based on 10-K business descriptions.

[24] Using lagged control variables and *tech differentiation* based on the patent portfolio from years $t-1$ to $t-5$ reduces concerns about reverse causality.

[25] Given the number of observations in our sample (especially in the split sample analyses later in the paper), we do not use industry-year- and technology(-year)-level fixed effects in our main specification because this would include several thousand additional fixed effects besides the several thousand fixed effects (firm level, industry level, year level) already included in our baseline specification. In addition, we do not include technology- and technology-year-level fixed effects in our main specification because these fixed effects represent an alternative means to characterize the technology portfolio of firms and to map firms' position in technology space based on the classifications of patents rather than patent text (e.g., Jaffe 1986). Including additional fixed effects based on patent classification runs counter to our effort to compare the use of patent text with patent classification to characterize the technology portfolio of firms. Nevertheless, as illustrated in Table A.8 in the online appendix, our main effects remain robust after including industry-year-level and/or technology- and technology-year-level fixed effects besides the firm-, industry-, and year-level fixed effects.

[26] We collect the disambiguated lawyer data from https://patentsview.org/download/data-download-tables and match each granted patent to its lawyer and law firm. On average, each patent portfolio is matched to 195 lawyers (median = 11) and eight law firms (median = 2).

[27] Columns (1) and (2) in Table A.15 in the online appendix show the corresponding regressions.

[28] Columns (3) and (4) in Table A.15 in the online appendix show the corresponding regressions.

[29] Table A.1 in the online appendix provides an explanation of how each measure is calculated. Data on product market competition are only available since 1989, so we lose nine years of observations in the regression analysis.

[30] As the only exception, *tech differentiation* does not have a significantly stronger relationship with Tobin's Q ($z = 0.933$, $p = 0.351$) if we split the sample based on the mean product market rivalry as measured using U.S. Census-based Herfindahl–Hirschman indexes.

References

Aghion P, Bloom N, Blundell R, Griffith R, Howitt P (2005) Competition and innovation: An inverted-U relationship. *Quart. J. Econom.* 120(2):701–728.

Ahuja G, Katila R (2001) Technological acquisitions and the innovation performance of acquiring firms: A longitudinal study. *Strategic Management J.* 22(3):197–220.

Ahuja G, Morris Lampert C (2001) Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic Management J.* 22(6–7):521–543.

Arinas I (2012) How vague can your patent be? Vagueness strategies in U.S. patents. *J. Language Comm. Bus.* 25(48):55–74.

Arora A, Gambardella A (1990) Complementarity and external linkages: The strategies of the large firms in biotechnology. *J. Indust. Econom.* 38(4):361–379.

Arora A, Belenzon S, Sheer L (2021) Matching patents to Compustat firms, 1980–2015: Dynamic reassignment, name changes, and ownership structures. *Res. Policy* 50(5):104217.

Arora A, Fosfuri A, Gambardella A (2001) Markets for technology and their implications for corporate strategy. *Indust. Corporate Change* 10(2):419–452.

Arts S, Cassiman B, Gomez JC (2018) Text matching to measure patent similarity. *Strategic Management J.* 39(1):62–84.

Arts S, Cassiman B, Hou J (2022) Technology differentiation, product market rivalry, and M&A transactions. Taneja S, ed. *Acad. Management Proc.* 2151–6561.

Arts S, Hou J, Gomez JC (2021) Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures. *Res. Policy* 50(2):104–144.

Bellstam G, Bhagat S, Cookson JA (2021) A text-based analysis of corporate innovation. *Management Sci.* 67(7):4004–4031.

Bloom N, Schankerman M, Van Reenen J (2013) Identifying technology spillovers and product market rivalry. *Econometrica* 81(4):1347–1393.

Bowen DE, Fresard L, Hoberg G (2023) Rapidly evolving technologies and startup exits. *Management Sci.* 69(2):940–967.

Carlson NA (2023) Differentiation in microenterprises. *Strategic Management J.* 44(5):1141–1167.

Cassiman B, Veugelers R (2002) R&D cooperation and spillovers: Some empirical evidence from Belgium. *Amer. Econom. Rev.* 92(4): 1169–1184.

Custódio C, Ferreira MA, Matos P (2019) Do general managerial skills spur innovation? *Management Sci.* 65(2):459–476.

Fama EF, French KR (1997) Industry costs of equity. *J. Financial Econom.* 43(2):153–193.

Gil Moltó MJ, Georgantzís N, Orts V (2005) Cooperative R&D with endogenous technology differentiation. *J. Econom. Management Strategy* 14(2):461–476.

Guzman J, Li A (2023) Measuring founding strategy. *Management Sci.* 69(1):101–118.

Hall BH, Jaffe AB, Trajtenberg M (2005) Market value and patent citations. *RAND J. Econom.* 36(1):16–38.

Hirshleifer D, Hsu P, Li D (2018) Innovative originality, profitability, and stock return. *Rev. Financial Stud.* 31(7):2553–2605.

Hoberg G, Philips G (2016) Text-based network industries and endogenous product differentiation. *J. Political Econom.* 124(5): 1423–1465.

Hotelling H (1929) Stability in competition. *Econom. J.* 39(153):41–57.

Igami M, Uetake K (2020) Mergers, innovation, and entry-exit dynamics: Consolidation of the hard disk drive industry, 1996–2016. *Rev. Econom. Stud.* 87(6):2672–2702.

Jaffe AB (1986) Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market value. *Amer. Econom. Rev.* 76(5):984–1001.

Jaffe AB (1989) Characterizing the "technological position" of firms, with application to quantifying technological opportunity and research spillovers. *Res. Policy* 18(2):87–97.

Jaffe AB, Trajtenberg M, Fogarty MS (2000) Knowledge spillovers and patent citations: Evidence from a survey of inventors. *Amer. Econom. Rev.* 90(2):215–218.

Jaffe AB, Trajtenberg M, Henderson R (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. *Quart. J. Econometrics* 108(3):577–598.

Kamien MI, Zang I (2000) Meet me halfway: Research joint ventures and absorptive capacity. *Internat. J. Indust. Organ.* 18(7):995–1012.

Kong N, Dulleck U, Jaffe AB, Sun S, Vajjala S (2023) Linguistic metrics for patent disclosure: Evidence from university vs. corporate patents. *Res. Policy* 52(2):104670.

Le Q, Mikolov T (2014) Distributed representations of sentences and documents. Xing P, Jebara T, eds. *Proc. 31st Internat. Conf. Machine Learn.*, 32:1188–1196.

Lin P, Zhou W (2013) The effects of competition on the R&D portfolios of multiproduct firms. *Internat. J. Indust. Organ.* 31(1): 83–91.

Mowery DC, Oxley JE, Silverman BS (1998) Technological overlap and interfirm cooperation: Implications for the resource-based view of the firm. *Res. Policy* 27(5):507–523.

Peteraf MA (1993) The cornerstones of competitive advantage: A resource-based view. *Strategic Management J.* 14(3):179–191.

Peters RH, Taylor LA (2017) Intangible capital and the investment-q relation. *J. Financial Econom.* 123(2):251–272.

Righi C, Simcoe T (2019) Patent examiner specialization. *Res. Policy* 48(1):137–148.

Rosenkopf L, Almeida P (2003) Overcoming local search through alliances and mobility. *Management Sci.* 49(6):751–766.

Rothaermel FT, Deeds DL (2004) Exploration and exploitation alliances in biotechnology: A system of new product development. *Strategic Management J.* 25(3):201–221.

Sampat BN (2018) A survey of empirical evidence on patents and innovation. National Bureau of Economic Research Working Paper No. 25383, National Bureau of Economic Research, Cambridge, MA.

Shaked A, Sutton J (1982) Relaxing price competition through product differentiation. *Rev. Econom. Stud.* 49(1):3–13.

Silverman BS (1999) Technological resources and the direction of corporate diversification: Toward an integration of the resource-based view and transaction cost economics. *Management Sci.* 45(8):1109–1124.

Simeth M, Cincera M (2016) Corporate science, innovation, and firm value. *Management Sci.* 62(7):1970–1981.

Stuart TE, Podolny JM (1996) Local search and the evolution of technological capabilities. *Strategic Management J.* 17(S1):21–38.

Sutton J (1991) *Sunk Costs and Market Structure* (MIT Press, Cambridge, MA).

Thompson P, Fox-Kean M (2005) Patent citations and the geography of knowledge spillovers: A reassessment. *Amer. Econom. Rev.* 96(1):450–460.

Wernerfelt B (1984) A resource-based view of the firm. *Strategic Management J.* 5(2):171–180.

Yayavaram S, Ahuja G (2008) Decomposability in knowledge structures and its impact on the usefulness of inventions and knowledge-base malleability. *Admin. Sci. Quart.* 53(2):333–362.