Research Policy 45 (2016) 81–96

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Mapping the technological landscape: Measuring technology distance, technological footprints, and technology evolution

Barak S. Aharonson[a,*], Melissa A. Schilling[b]

[a] Tel Aviv University, Business School, Strategic Management, 6997801 Tel Aviv, Israel
[b] New York University, Stern School of Business, 44 West 4th Street, New York, NY 10012 USA

ARTICLE INFO

Article history:
Received 29 April 2014
Received in revised form 23 July 2015
Accepted 3 August 2015

Keywords:
Technology landscape
Knowledge-spillover
Recombinant search

ABSTRACT

We develop and apply a set of measures that enable a fine-grained characterization of technological capabilities based on the USPTO database. These measures can capture the distance between any two patents, and help to identify outlier patents. They also provide a rich characterization of a firm's technological footprint, including its depth and breadth. The measures also enable researchers to assess the technological overlap, similarity, and proximity of the technological footprints of two or more firms. At the level of the macro technology landscape, the measures can be used to explore such dynamics as technology agglomeration, knowledge spillovers, and technology landscape evolution. We show applications of each of the measures and compare the results obtained with those that would be obtained with previously existing measures of firm diversity, similarity and proximity, highlighting the advantages of the measures used here.

1. Introduction

A firm's technological capabilities are central to its identity, its strategies, and its potential for success. Technological capabilities represent what the firm can do in the present, as well as what it has learned in the past. These capabilities also have a significant influence on the trajectories a firm will choose in the future. It is not surprising, then, that there are many streams of research that invoke the notion of technological capabilities, including (but not limited to) those that examine competitive positioning, innovation, organizational learning, diversification, and organizational growth. However, the development of measures for constructs such as technological capabilities, technology distance, technological similarity, and technology footprints is a nascent field, with much opportunity for further development.

Most of the prior work on such measures can be classified into four categories: (1) measured based on SIC codes, (2) measures based on patent classes, (3) measures based on patent citations, and (4) measures based on textual analysis of patents. Early work on technological relatedness tended to use industry classification (SIC or NAICS) codes, looking at, for example, a firm's business share across different SIC codes or the degree to which its combinations of SIC codes mirrored patterns in the general economy – what Teece et al. (1994) term "corporate coherence." The advantage of this kind of approach is that one is not limited to examining only firms that patent. The main disadvantage of this kind of approach is its imprecision: First, most industry classification systems reflect markets in which a firm competes, not technological capabilities with which they compete. For example, SIC code 2043, "cereal breakfast foods manufacturer," tells us little about the technology (e.g., cooking, drying, extrusion, rolling, etc.) involved. By contrast, USPC mainline subclass 323.4 "cereal puffing by means of an apparatus adapted to subject cereal to sudden changes in pressure to disrupt the same and produce an expanded or inflated product" gives us considerably more technology capability-relevant information and offers greater precision. Second, SIC codes are typically only available at the firm level, and even then usually only at the "primary" SIC code level, which may mask considerable firm activity.

One of the most common approaches to handling technology distance or similarity is to count the number of patent classes two firms have in common across their patenting portfolios (e.g., Ahuja and Katila, 2001; Diestre and Rajagopalan, 2012; Dushnitsky and Lenox, 2005). This method is relatively easy to implement and relies on publicly held data that can be regularly updated, however the tradeoff for this efficiency is a loss of granularity and detail. This loss is particularly acute when researchers use only the first class listed on a patent (Benner and Waldfogel, 2008). As we will show, measures based on patent subclasses provide a much richer and more reliable picture of a firm's technological capabilities.

* Corresponding author.
E-mail addresses: aharonson@tau.ac.il (B.S. Aharonson), mschilli@stern.nyu.edu (M.A. Schilling).

http://dx.doi.org/10.1016/j.respol.2015.08.001
0048-7333/

There is also a significant body of work that measured technological relatedness by looking at whether firms (or patents) cite the same prior art. Some studies have recently highlighted challenges with using patent citations for measures of relatedness, however, because the citation of prior art is both discretionary and strategic, and the citation process was never designed to represent a taxonomy (e.g., Alcacer and Gittelman, 2006; Benner and Waldfogel, 2008). Furthermore, the patent citation process necessarily suffers from problems related to sequential interdependence: one can only cite work that has been made public *previously*, thus if patents are not assigned prior art citations at the same time, they draw from different opportunity sets of prior art.

Nascent work on textual analysis of patents by authors such as Gerken and Moehrle (2012), Yoon and Kim (2012) and Yoon et al. (2013) use the words and sentence structures in the patents to create a map of the technological positions and differences between patents. This is a very promising avenue of research that has the potential to yield rich maps of firms' technology capabilities. However, to create these measures, a researcher typically must understand the technological domains and their terms, synonyms and acronyms. As a result, this approach is usually deployed in a single narrow domain rather than across multiple technology domains.

Precise and fine-grained measures of technological capabilities might enable researchers to better characterize such features as a firm's technological footprint, how broad or specialized that footprint is, how distant its technological capabilities are from those of its peers (or how distant the technological capabilities of an alliance partner or acquisition target are from those of a given firm), how much the firm's technological footprint overlaps with those of competitors, etc. Such measures could also be used to better understand dynamic processes such as how a firm has developed over time, how diversification moves build upon or extend technological capabilities, or even the evolution of entire technological fields.

We thus develop and test here a set of measures that enable a finer-grained characterization of technological capabilities. These measures enable researchers to create a rich characterization of a firm's technological footprint, including its depth and breadth. The measures also enable researchers to assess the technological overlap, similarity, and proximity of two or more firms. The measures make it straightforward to see how a firm's moves relate to its existing technological capabilities, such as whether an acquisition enhances an existing technological strength or represents an extension into related or unrelated fields. At a macro level, the measures can be used to explore knowledge spillovers, such as how the existence of intellectual capital about a particular set of technologies influences the growth of intellectual capital in adjacent technology positions. Our approach also provides a method of identifying and demarcating industry boundaries in a way that is not always clearly captured by conventional industry classification systems. This may be particularly useful for understanding the emergence of (and relationships between) nascent industries.

We begin by explaining the notion of recombinant search on a technology landscape, and then develop a set of measures of technology positions on that landscape that permit us to create fine-grained measures of technological distance and technology footprints. Though the concept of a technology landscape underlies much of the work on recombinant search and innovation (e.g., Ahuja and Katila, 2004; Fleming and Sorenson, 2001; Fleming and Sorenson, 2004; Kauffman et al., 2000; Rosenkopf and Almeida, 2003), only a few studies have attempted to examine what the actual technology landscape looks like or how technology positions grow within it (e.g., Podolny and Stuart, 1995; Stuart and Podolny, 1996; Schoen et al., 2012). Like their work we utilize patent data to characterize the technology landscape; unlike their work, however, we do not rely on patent citations to represent the knowledge a firm has or builds upon, nor are we constrained to a particular industry. To develop our measures, we first use a novel technology distance measure to map the entire database of USPTO patents filed from 1976 to 2012, identifying every occupied technology position and its distance from other occupied technology positions. We then show how those technology positions can be used to calculate technological distance, and to assemble technology footprints of firms. These measures, in turn, can be used to assess such dimensions as technological concentration, technological similarity, technological proximity, and other useful measures that can be applied at the patent level, technology position level, firm levels, or population levels.

2. Recombinant search on a technology landscape

The world of potential technological innovations can be conceived of as a landscape, with each potential position on the landscape corresponding to a particular configuration of components (Fleming and Sorenson, 2004; Kauffman et al., 2000). Each position may or may not be occupied by existing innovations; some positions are more feasible or valuable than others, and some positions are potentially feasible or valuable but innovations have not yet been discovered in them. Furthermore, the distribution of innovations across this technology space is unlikely to be either random or uniform. Instead, for reasons we will discuss in the paper, innovations are likely to agglomerate into clusters of adjacent technology positions.

The process of exploring different potential solutions to a problem, including the identification of new knowledge elements or new relationships between knowledge elements, is often termed recombinant search. *Recombinant* refers to the fact that the creation of new knowledge is most often the result of novel recombinations of known elements of knowledge, problems, or solutions (Gilfillan, 1935; Nelson and Winter, 1982; Penrose, 1959; Schumpeter, 1934; Usher, 1954) or the reconfiguration of the ways in which such knowledge elements are linked (Henderson and Clark, 1990; Henderson, 1994). Nelson and Winter (1982: 130) provide a cogent summary of this argument: "innovation in the economic system – and indeed the creation of any sort of novelty in art, science, or practical life – consists to a substantial extent of a recombination of conceptual and physical materials that were previously in existence."

2.1. Technological distance and agglomerations

Much of the research on recombinant search implicitly or explicitly utilizes a spatial metaphor. Reference is made to searching in the "neighborhood" of past solutions (Cyert and March, 1963) or practicing "local" or "distant" search (Nelson and Winter, 1982). Such a conception of search invokes the notion of a technology landscape, characterized by peaks of opportunity wherein particular combinations of knowledge elements yield valuable new solutions, and valleys wherein particular combinations of knowledge components are unlikely to be useful. Most of this technology landscape is unknown to individual inventors or firms, leading firms to focus on only the familiar portions of the technology landscape, i.e., the technological elements (and combinations of elements) with which they have prior experience (Sorenson and Fleming, 2004; Stuart and Podolny, 1996). This is one of the primary explanations for the prevalence of local search. The more knowledge and experience an inventor has in a particular domain, the more likely they are to understand the nature of the relationships between different technological elements, and the more efficient the inventor should be in searching for a useful combination (Cohen and Levinthal, 1990; Dosi, 1988; Harlow, 1959). Furthermore, to the degree that