Economics Letters 116 (2012) 457–459

Contents lists available at SciVerse ScienceDirect

Economics Letters

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

# A measure of technological distance[☆]

Talia Bar[a,*], Aija Leiponen[b,c]

[a] Department of Economics, Cornell University, United States
[b] Dyson School of Applied Economics and Management, Cornell University, United States
[c] Imperial College London, United Kingdom

ARTICLE INFO

*Article history:*
Received 14 May 2009
Received in revised form
25 February 2012
Accepted 16 April 2012
Available online 23 April 2012

*JEL classification:*
O30

*Keywords:*
Patents
Research and development
Technological distance

ABSTRACT

In this paper we construct an intuitive measure of technological distance. We compare it to previously used measures and show that it satisfies a desirable independence axiom that other commonly used measures fail to satisfy.



## 1. Introduction

In empirical research in areas such as economics of growth, technical change, and innovation it is often important to measure the technological distance between firms or industries. Technological distance measures have been used for example by Jaffe (1986, 1989) in the context of technology spillovers, Stuart and Podolny (1996) in their work on geographic proximity and its effects on corporate R&D, Gilsing et al. (2008) and Rosenkopf and Almeida (2003) on alliances, and Bar and Leiponen (2012) on cooperative standard setting. Typically, patent data is used to characterize firms' technological positions. Research and development distance (or proximity) between firms is measured by comparing vectors that represent firms' shares of patents in each patent class. Previously used measures include the Euclidean distance between these vectors, the angle between them, or their correlation (see Benner and Waldfogel, 2007).

We propose to use a measure of technology distance that we refer to as the *min-complement* distance. We show that this measure is equivalent to the $L_1$ metric when defined on probability vectors. The way we formulate the measure lends itself to an intuitive interpretation in our context—it measures the share of non-overlapping inventions in the patent portfolios of two firms. Additionally, we argue that the min-complement distance measure satisfies a desired property that we call *Independence of Irrelevant Patent Classes*: a focal firm's distance from other firms only depends on relevant patent classes. Relevant classes include those in which the focal firm has some patent holdings. Hence, a firm's distance from other firms does not depend on the other firms' allocation of patents in classes that are empty for the focal firm. This property is not met by other commonly used technology distance measures.

## 2. Min-complement distance

We begin by proposing a measure of distance – the *min-complement distance* – which is defined on the finite $n$-dimensional simplex,

$$S = \left\{ P = (p_1, \ldots, p_n) \mid p_k \geq 0 \text{ and } \sum_{k=1}^{n} p_k = 1 \right\}.$$

**Definition 1** (*The Min-Complement Distance Measure*)**.** For any $P_i = (p_{i1}, \ldots, p_{ik}, \ldots, p_{in})$ and $P_j = (p_{j1}, \ldots, p_{jk}, \ldots, p_{jn})$ in $S$ let

$$M(P_i, P_j) = 1 - \sum_{k=1}^{n} \min\{p_{ik}, p_{jk}\}. \quad (1)$$

[☆] We are grateful to Haim Bar, Mukul Majumdar and an anonymous referee for their helpful suggestions.

* Correspondence to: Economics Department, Cornell University, Uris Hall, Ithaca, NY 14853, United States. Tel.: +1 607 255 6209.
*E-mail address:* tb97@cornell.edu (T. Bar).

0165-1765/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2012.04.030

The min-complement distance measure takes values $M \in [0, 1]$, with $M = 0$ being the closest distance. To illustrate the definition of our measure consider a simple example.

**Example 1.** If $P_1 = (\frac{1}{2}, \frac{3}{8}, \frac{1}{8})$ and $P_2 = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ then, $M(P_1, P_2) = 1 - \frac{1}{4} - \frac{3}{8} - \frac{1}{8} = \frac{1}{4}$.

In Proposition 1 we show that when defined on the set of probability vectors $S$, the min-complement distance measure is equivalent to the well known $L_1$ metric

$$L_1(P_1, P_2) = \sum_k |p_{1k} - p_{2k}|.$$

This implies that on the set $S$ of probability vectors the measure $M(P_i, P_j)$ is a well-defined distance measure. That is, the following properties hold for all $P_1, P_2, P_3 \in S$

1. (non-negative) $M(P_1, P_2) \geq 0$ and $M(P_1, P_2) = 0$ if and only if $P_1 = P_2$
2. (symmetric) $M(P_1, P_2) = M(P_2, P_1)$
3. (triangle inequality) $M(P_1, P_2) + M(P_2, P_3) \geq M(P_1, P_3)$

**Proposition 1.** *On* $S$, $M(P_1, P_2) = \frac{1}{2}L_1(P_1, P_2)$.

The proof is provided in Appendix. Proposition 1 immediately implies that the min-complement measure satisfies the distance properties because $L_1$ satisfies them.

## 3. Comparison with commonly used technology distance measures

### 3.1. Symmetric measures

In applied research, technological differences between firms have often been measured as the distance between their patent portfolios (see for example Benner and Waldfogel, 2007). A patent portfolio vector $P$ specifies for each firm the share of its patents in every one of $n$ relevant patent classes. Let $p_{ik}$ be the share firm $i$ has in class $k$. Three measures have typically been used:

The **Euclidean distance**:

$$E(P_i, P_j) = \sqrt{\sum_k (p_{ik} - p_{jk})^2}.$$

$E \in [0, 1]$ with $E = 0$ being the closest.

The **Angle** between firms $i$ and $j$ (which is in fact a measure of proximity) is the *cosine* of the angle between the two vectors $P_i$ and $P_j$:

$$A(P_i, P_j) = \frac{\sum_k p_{ik}p_{jk}}{\sqrt{\left(\sum_k p_{ik}^2\right)\left(\sum_k p_{jk}^2\right)}}.$$

$A \in [0, 1]$ with $A = 1$ being the closest.

The **Correlation** between portfolios $P_i$ and $P_j$:

$$C(P_i, P_j) = \frac{cov(P_i, P_j)}{\sqrt{var(P_i)var(P_j)}}.$$

$C \in [-1, 1]$ with $C = 1$ being the most positively correlated, and $C = (-1)$ the most negatively correlated and $C = 0$ for uncorrelated.

We propose the min-complement distance between patent portfolios, defined in (1), as an alternative measure of technology distance. Intuitively, in computing technological distance using the min-complement measure we subtract from the unit (capturing the whole universe of a firm's inventions) the share of invention overlap. In each patent class $k$ the overlapping share is $\min\{p_{ik}, p_{jk}\}$. Subtracting the sum of these shares of overlap from 1 we are left with a share of non-overlapping inventions, which we interpret as the technological distance.

We claim that the min-complement distance measure, $M()$, has advantages over the existing measures. We have already shown in the previous section that $M()$ is a well-defined distance measure. In contrast, the Angle and the Correlation measures are not. Moreover, we show that the min-complement measure satisfies an "*Independence of Irrelevant Patent Classes*" (IIPC) property which we define next. We later show that none of the other technology distance measures we listed above satisfies this property.

Consider three firms with patent portfolios $P_0, P_1$ and $P_2$. The IIPC property states that if the portfolios $P_1$ and $P_2$ are the same in every class which is relevant in portfolio $P_0$ (in patent classes in which $P_0$ has a strictly positive share), then their distance from $P_0$ should be the same. We formally define it as follows:

**Definition 2** (*Independence of Irrelevant Patent Classes*)**.** A distance measure $d$ satisfies IIPC if for any R&D portfolio vectors $P_0, P_1$ and $P_2$ such that for all $k$, $P_{0k}(P_{2k} - P_{1k}) = 0$, we have $d(P_0, P_1) = d(P_0, P_2)$.

For portfolio $P_0$ the relevant patent classes are those with $P_{0k} > 0$. The condition "for all $k, P_{0k}(P_{2k} - P_{1k}) = 0$ " means that the portfolios $P_1$ and $P_2$ are the same in all patent classes that are relevant for portfolio $P_0$, (those with $P_{0k} > 0$) and may differ only in irrelevant classes (where $P_{0k} = 0$). A distance measure satisfies independence of irrelevant patent classes property when the distance between portfolio $P_0$ and each of the two technology profiles $P_1, P_2$ is the same whenever these two portfolios differ from each other only in irrelevant patent classes, but have equal shares in all relevant patent classes. We argue that IIPC is a desired property for technology measures. The technology distance between two firms should depend on the shares of patents these firms have in classes in which they both actively patent, but should not depend on how patents for one firm are distributed between classes in which the other firm does not patent at all.[1] Proposition 2 shows that the min-complement distance measure satisfies the Independence of Irrelevant Patent Classes property.

**Proposition 2.** *The measure $M(., .)$ satisfies Independence of Irrelevant Patent Classes.*

In the following example we illustrate this property of our distance measure and we demonstrate that other commonly used distance measures do not satisfy it.

**Example 2.** Consider the patent portfolio $P_0 = (1, 0, 0)$ and the family of patent portfolios $P_\varepsilon = (a, \varepsilon, 1 - a - \varepsilon)$ where $1 > 1 - a > \varepsilon \geq 0$. Indeed $M(P_0, P_\varepsilon) = 1 - a$ (since the share of overlap is $a$) which does not depend on $\varepsilon$. Since portfolio $P_0$ has no patent holdings in classes 2 and 3, its distance from any portfolio $P_\varepsilon$ should not depend on how the share of patents is allocated between these two irrelevant classes. Note that all three previously proposed distance measures violate IIPC since $A, C$ and $E$ vary with $\varepsilon$:

$$E(P_0, P_\varepsilon) = \sqrt{(1-a)^2 + \varepsilon^2 + (1-a-\varepsilon)^2}.$$

$$A(P_0, P_\varepsilon) = \frac{a}{\sqrt{(a^2 + \varepsilon^2 + (1-a-\varepsilon)^2)}}$$

$$C(P_0, P_\varepsilon) = \frac{a - \frac{1}{3}}{\sqrt{\frac{2}{3}\left[\left(a - \frac{1}{3}\right)^2 + \left(\varepsilon - \frac{1}{3}\right)^2 + \left(\frac{2}{3} - a - \varepsilon\right)^2\right]}}$$

[1] We note, however, that sometimes firms who have patents in different but related patent classes could be technologically close. Depending on the application and classes used, one might want to count such technologically related classes as belonging to the same "bin", and include them in the same component of the patent portfolio vector $P$.

The min-complement measure of technological distance is useful for empirical research on innovation networks, spillovers, technological competition, and industrial evolution. We believe that its Independence of Irrelevant Patent Classes property is an important attribute. We argue that the potential for technological spillovers, R&D competition, or collaboration between a firm with some empty technology classes and other firms should depend on the other firms' patent holdings in the technological areas where the firm has intellectual property (the relevant classes), and not on how the other firms' technological assets are distributed in areas that are not relevant to the focal firm. This is particularly important because small innovative firms are less likely to have very broad portfolios. Hence they will have many empty technology classes, whereas technology giants tend to have patents in dozens of classes. The min-complement distance of a focal firm from other firms depends only on the shares of patents these firms have in the areas where the focal firm is active. However, take for example the Euclidian distance, for two firms with equal shares in the relevant patent classes; a small firm (whose patents in the irrelevant areas are concentrated in a few classes) will appear to be more distant from the focal firm than a large firm who holds patents in many of the irrelevant areas, even when these two firms have exactly the same shares in the relevant classes.

### 3.2. Symmetric vs. asymmetric measures

The min-complement measure satisfies the mathematical properties of a distance (or metric); in particular, it is symmetric. Symmetry is a standard and intuitive assumption to make on distance measures. It holds true for commonly used technology distance measures with which we compare the measure proposed in this paper. It seems appropriate to use a symmetric distance measure for example when the purpose is to identify how closely related the R&D activities of two companies are, or how substitutable their R&D capabilities are. A short distance could indicate the potential for R&D competition but also for mutual learning and R&D collaboration. However, for some applications an asymmetric measure is more appropriate. For example, Mowery et al. (1996) examine interfirm knowledge transfers in alliances. They use an asymmetric index based on cross-patent citations to measure the degree to which one firm acquires technology-based capabilities from another. In a study of knowledge flows between European regions, Maurseth and Verspagen (2002) propose an asymmetric measure of compatibility based on patent citations. Asymmetry might be more reasonable when it is meant to identify the extent to which one company builds on the knowledge of another. Put in other words, symmetric measures can identify a horizontal distance, while asymmetric measures can better capture also the potential for vertical spillovers, and directional information flows.

The min-complement measure, together with the other symmetric measures with which we compared it, are all defined on patent portfolio shares. Thus, for given shares, these measures are independent of firm size. According to these distance measures, firms that own intellectual property rights in the same classes are considered close even if they own portfolios of different sizes. It is natural that a symmetric distance measure would not depend on firm size — the distance measure from a large to a small firm would equal that from a small to a large firm. If the researcher prefers a measure that depends on firm size, there may be need to employ an asymmetric measure.

## 4. Conclusions

Measuring technological distance between firms based on their patent holdings is useful for empirical studies of inventive activity. We propose the min-complement distance which is defined on vectors of shares of patents in patent classes. This measure is highly correlated with commonly used measures but has some advantages over them. It has an intuitive interpretation and it satisfies the Independence of Irrelevant Patent Classes property. That is, the distance between a firm and other firms does not depend on the firms' distribution of patents between classes which are not relevant for the focal firm.

## Appendix

**Proof** (Proposition 1)**.** We show that $M(P_1, P_2) = \frac{1}{2}L_1(P_1, P_2)$. Let $A$ be the set of indices for which $p_{1k} \le p_{2k}$.

$$\begin{aligned} M(P_1,P_2) &= 1 - \sum_k \min\{p_{ik}, p_{jk}\} = 1 - \sum_{k\in A} p_{1k} - \sum_{k\notin A} p_{2k} \\ &= \frac{1}{2}\left(\sum_k p_{1k} + \sum_k p_{2k} - 2\sum_{k\in A} p_{1k} - 2\sum_{k\notin A} p_{2k}\right) \\ &= \frac{1}{2}\left(\sum_{k\notin A} p_{1k} + \sum_{k\in A} p_{2k} - \sum_{k\in A} p_{1k} - \sum_{k\notin A} p_{2k}\right) \\ &= \frac{1}{2}\left(\sum_{k\notin A} (p_{1k} - p_{2k}) + \sum_{k\in A} (p_{2k} - p_{1k})\right) \\ &= \frac{1}{2}\sum_k |p_{1k} - p_{2k}| = \frac{1}{2}L_1(P_1, P_2). \end{aligned}$$

It immediately follows that $M(.,.)$ satisfies the properties of a distance measure 1–3. □

**Proof** (Proposition 2)**.** Suppose $P_0, P_1$ and $P_2$ are portfolio vectors such that for all $k$, $P_{0k}(P_{2k} - P_{1k}) = 0$,

$$\begin{aligned} M(P_0,P_1) &= 1 - \sum_k \min\{p_{0k}, p_{1k}\} \\ &= 1 - \sum_{\{k:p_{0k}>0\}} \min\{p_{0k}, p_{1k}\} - \underbrace{\sum_{\{k:p_{0k}=0\}} \min\{p_{0k}, p_{1k}\}}_{=0} \\ &= 1 - \sum_{\{k:p_{0k}>0\}} \min\{p_{0k}, p_{2k}\} - \underbrace{\sum_{\{k:p_{0k}=0\}} \min\{p_{0k}, p_{2k}\}}_{=0} \\ &= 1 - \sum_k \min\{p_{0k}, p_{2k}\} = M(P_0, P_2) \quad \square \end{aligned}$$

## References

Bar, T., Leiponen, A., 2012. Collaboration and Networking in Cooperative Standard Setting. Cornell University.
Benner, M., Waldfogel, J., 2007. Close To You? Bias and Pre in Patent-Based measures of Technological Proximity. NBER Working Paper 13322, http://www.nber.org/papers/w13322.
Gilsing, V., Nooteboom, B., Vanhaverbeke, W., Duysters, G., van den Oord, A., 2008. Network embeddedness and the exploration of novel technologies: technological distance, betweenness centrality and density. Research Policy 37, 1717–1731.
Jaffe, A., 1986. Technological opportunity and spillovers of R&D: evidence from firms' patents, profits, and market value. American Economic Review 76, 985–1001.
Jaffe, A., 1989. Characterizing the 'technological position' of firms, with application to quantifying technological opportunity and research spillovers. Research Policy 18, 87–97.
Maurseth, P.B., Verspagen, B., 2002. Knowledge spillovers in Europe: a patent citations analysis. Scandinavian Journal of Economics 104 (4), 531–545.
Mowery, D.C., Oxley, J.E., Silverman,, Brian, S., 1996. Strategic alliances and interfirm knowledge transfer. Strategic Management Journal 17, Issue Special Issue: Knowledge and the Firm, 77–91.
Rosenkopf, L., Almeida, P., 2003. Overcoming local search through alliances and mobility. Management Science 49 (6), 751–766.
Stuart, T.E., Podolny, J.M., 1996. Local search and the evolution of technological capabilities. Strategic Management Journal 17, 21–38.