ISSN: 2281-1346

Department of Economics and Management

DEM Working Paper Series

Cognitive distance in research collaborations

Margherita Balconi
(Department of Economics and Management, University of Pavia)

Valeria Lorenzi
(Dipartimento di Scienze Economico Aziendali, Università di Milano Bicocca, Italy)

Pier Paolo Saviotti
(INRA GAEL, Univérsité Pierre Mendés-France, Grenoble, France)

Antonella Zucchella
(Department of Economics and Management, University of Pavia)

 # 51 (09-13)

Via San Felice, 5
I-27100 Pavia
http://epmq.unipv.eu/site/home.html

September 2013
Cognitive distance in research collaborations

M. Balconi°, V.Lorenzi^, P.P. Saviotti*, A. Zucchella[#]

° Dipartimento di Scienze Economiche e Aziendali, Università di Pavia, Italy;
address: via San Felice 5, I27100 Italy;
tel. +39 0382 986225; fax: +39 0382 422583
e-mail: balconi@unipv.it

^ Dipartimento di Scienze Economico Aziendali, Università di Milano Bicocca, Italy;
e-mail : valeria.lorenzi85@gmail.com

*INRA GAEL, Univérsité Pierre Mendés-France, Grenoble, France;
e-mail: pier-paolo.saviotti@wanadoo.fr

[#] Dipartimento di Scienze Economiche e Aziendali, Università di Pavia, Italy;
e-mail: antonella.zucchella@unipv.it

Abstract

This paper addresses the cognitive dimension of proximity/distance in research collaborations of small biotechnology firms. While the theory of optimal cognitive distance assumes learning as motive of collaborations, we suggest that small specialised firms or sub-units of big diversified organizations tend to collaborate with actors endowed with different specific knowledge, with the purpose of accessing rather than acquiring and assimilating the pieces of knowledge they lack. This leads to the expectation of high cognitive distance between collaborators, who however can understand each other since they share the same basic knowledge.

We apply this framework to investigate the research collaborations of a sample of Italian biotech firms, using data on the papers co-authored by individuals belonging to different organizations. In order to measure cognitive distance at a very disaggregated level, we introduce an index originally developed by ecologists to measure distance between different species. As expected, most co-publishing partners have high cognitive distance. Moreover, the knowledge accessing motive is also consistent with the finding that even small firms often engage in extensive networks of collaborations with a remarkable variety of actors and do not tend to enlarge their scope over time.

Keywords: biotech; cognitive distance; knowledge; research collaborations.

1. Introduction

Firms engaged in complex and uncertain research and innovation processes depend crucially on interactions and collaborations with external actors, in order to overcome the limitations of specific knowledge bases. This is particularly manifest in research-intensive sectors like biotechnology, where the underlying science is very dynamic and the technologies are strongly science-based (Powell et al., 1996; Pisano, 2007). Depending on the desired balance between exploration and exploitation, two main kinds of network are found in biotechnology (March, 1991; Gilsing and Nooteboom, 2005; Rothermael and Deeds, 2004; Stuart et al., 2007): collaborations between universities, research centres and dedicated biotechnology firms (DBFs), based mostly on the exploration and creation of knowledge; and partnerships between DBFs and large pharmaceutical companies, based primarily on the exploitation of knowledge and commercialization of innovations.

This paper focuses exclusively on research collaborations. On this topic Liebeskind et al. (1996) found that in order to source scientific knowledge biotechnology firms attribute more importance to *individual-level ties* between their researchers and university scientists than to interorganizational agreements. Henderson and Cockburn (1994) showed that for pharmaceutical firms an important source of strategic advantage is the ability to access and integrate knowledge not only across the boundaries of firms, but also *across disciplines and therapeutic class boundaries*. Cockburn and Henderson (1998) also stressed that in order to access the results of publicly funded research, firms' researchers need *to be active participants in the construction of publicly available research results* (that is, they need to publish), despite the issues of appropriability that such active collaboration raises. According to Muller and Pénin (2006), publishing in scientific journals helps firms to develop R&D collaborations with other firms or public institutions, by improving their external reputation.

In turn R&D cooperation allows an efficient division of research labor, since it enables firms to develop synergies with complementary partners. That biotech and pharmaceutical companies publish heavily has been widely documented, with annual counts of papers comparable to, and sometimes exceeding, the output of similarly sized universities and research institutions (Hicks, 1995). Comparing the data on scientific publications co-authored by firms' researchers and researchers of other organizations with data on strategic alliances, Gittelman (2007) found that co-publications capture a substantial proportion of collaborative research partnerships but give a much wider picture of the research networks of the firms than alliance data.

Drawing upon this stream of literature, we examine the interactions of dedicated biotechnology firms (DBFs) with other organizations, such as universities, hospitals and other firms, which are revealed by coauthorships of scientific papers across institutions (Cockburn and Henderson, 1998). The high number of interactions per firm we find is likely to reflect for the most part informal collaborations enacted by individual researches across organizational boundaries, rather than formal alliances (about which we do not have data). In this context, the main objective of the paper is to study the relationship between the cognitive distance (CD) of organizations and the probability of collaborations, where CD is defined as the degree of overlapping between the specialised knowledge bases of the organizations involved in a collaboration (Nesta and Saviotti, 2005; Pyka and Saviotti, 2005). In order to formulate our expectations, we first need to provide a new conceptual framework, placing the issue of cognitive distance within the logic or research collaborations, be they formal or mostly informal. Moreover, in order to test our expectations, we introduce a new methodology to measure CD suited to capture the similarity/dissimilarity of the knowledge profiles of specialized organizations, using a very disaggregated classification of knowledge fields.

We think a new framework is needed, since the approach developed by the (rather scarce) existing literature on cognitive distance is not adequate to interpret a context of mostly informal knowledge exchanges established by small specialized firms through the activity of border-crossing researchers. This literature is based on the hypothesis that the main purpose pursued by collaborating firms is *mutual learning* (Nooteboom, 2000 & 2007; Wuyts et al., 2005).

We argue that assuming learning as the overwhelming motivation driving firms to collaborate is in contrast both with the fact that small firms have a multiplicity of links with different sources of knowledge and with the general trend observed in the industrial landscape where most specialised actors do no tend to enlarge the scope of their activity and become more similar to one another, as would happen if they actively sought to mutually learn. Following Mowery et al. (1996) and Grant and Baden-Fuller (2004), we propose the distinction between the motive of learning and that of accessing the partners' knowledge, without attempting to assimilate, acquire and apply it autonomously for commercial ends. In fact accessing and combining different pieces of knowledge, research instruments and experimentations through interactions, rather than learning, is consistent with enduring firm specialization and focus. Only some of these interactions, if any, might set the ground for successive learning, which involves allocating human and material resources in new directions. Moreover, since the real knowledge carriers are the individuals who are at the same time organizations' employees and members of a work-based epistemic collectivity that transcends organizational borders, recognizing the importance of this overarching community is crucial[1]. In fact it is within the basic knowledge space delimiting an epistemic community

[1] It is worth noting that a company's knowledge base comprises both the (tacit and codified) knowledge possessed by its individual employees and used during their work activity and the knowledge embedded in

that the specialised knowledge of its individual members flows and is effectively exchanged. Accordingly, a very disaggregated classification of knowledge fields is required in order to capture the nature of these flows.

In the second part of the paper we apply our interpretative framework to the research collaborations revealed by co-publications of a sample of 31 Italian biotechnology firms. From the analysis of the 511 scientific papers co-authored with other organizations, 1.244 collaborations emerged with 845 other entities (other firms, hospitals, universities and research organizations). Since many of the collaborating organizations are big and have a wide scope, in order to measure cognitive distance we consider only the subunits directly involved. More precisely, the paper is organised as follows. In section 2 we discuss the literature and construct the conceptual framework. In section 3 we explain the methodology used to measure cognitive distance. Section 4 describes the data and the classifications adopted. Section 5 illustrates the structure of collaborations among the various types of actors involved (firms, universities, hospitals, research organizations) while section 6 shows and discusses the results obtained with regard to cognitive distance. Section 7 draws some conclusions and proposes suggestions for future research.

2. Conceptual framework

What do we know about cognitive distance/proximity between collaborating partners? The vast literature on alliances, networks and clusters (Breschi and Malerba, 2005) has devoted a great deal of attention to the issue of geographic proximity, but a very scant interest to the cognitive dimension of distance. There is wide consensus that geographic proximity

---

organizational and managerial routines (Tijssen 2004). However, in the case of companies engaged only in R&D with no production activity, the weight of the scientific and technical knowledge of its teams of researchers is overwhelming.

matters especially for the communication of tacit knowledge, which necessarily involves face to face contacts, while codified knowledge may also be accessed at distance (Nonaka, 1994; Nonaka and Takeuki, 1995; Storper, 1997). Face to face contacts also help create trust, while the sharing of experiences and technologies as well as the mobility of employees enhance learning processes (Feldman, 2000; Cooke, 2002).

The concept of cognitive distance - or the converse, cognitive proximity - has been proposed by Nooteboom (2000, Nooteboom et al. 2007) and defined as a way to interpret resource heterogeneity between the firms involved in alliances. Nooteboom develops a theory of the “optimal” degree of CD, which is based on the assumption that *mutual learning* is the objective driving different actors to interact. In inter-firm relationships learning entails a trade-off between the advantage of high cognitive distance in terms of novelty value of the partner’s knowledge, and the disadvantage of low mutual understanding due to the problem of absorptive capacity. This notion, according to Cohen and Levinthal (1990), comprises the abilities to recognize the value of new information, assimilate it, and apply it to commercial ends. Thus a firm can expect a greater extent of learning from an organization having the knowledge it wants to acquire than its same knowledge. However, the costs and difficulties of communication and assimilation tend to rise with a growing cognitive distance, till mutual understanding is precluded. “If effectiveness of learning by interaction is the mathematical product of novelty value and understandability, the result is an inverse U-shaped relation with cognitive distance” (Wuyts et al., 2005, p. 279; Nooteboom et al., 2007; Gilsing et al., 2008). Thus the optimal CD is the degree of CD which produces the most effective learning and hence the best innovative performance. To test the main hypothesises derived from the theory, this literature empirically examines the formal technological alliances between the largest companies in some industries (the chemical, automotive, pharmaceutical and ICT industries) and classifies the various fields of knowledge mastered by collaborating firms at a very high level of aggregation (on this aspect, see the next section).

While this theoretical proposal can help understand the choice of alliance partners made by big firms aiming at enlarging the scope of their activities, it does not appear suited to be applied to the empirical landscape investigated in this paper. First, our focus is on the myriad of mostly informal research collaborations the importance of which as a source of scientific knowledge has already been widely recognised by the literature (Powell et al.1996; Liebeskind et al. 1996; Oliver and Liebeskind, 1998; Gittelman 2007). In particular we study the research collaborations which small specialised firms, engaged almost exclusively in R&D, realize with a number of different actors, mainly with the aims of solving the problems which they are not able to solve alone (lacking the required resources or knowledge, Liebeskind et al., 1996) or of undertaking research projects that need the pooling of competences and resources. The disciplines, skills and instrumentation specialist actors need to reach are only occasionally useful (Feldman, 1994) and it would not pay investing to acquire them. Moreover, ties with university scientists allows DBFs to access the latest development in research and instrumentation (Oliver and Liebeskind, 1998). Some of these collaborations may also be "prospecting," allowing the firms to "peek" at others' knowledge at low cost (Liebeskind et al., 1996).

Accessing the required capabilities is mostly accomplished through the direct initiatives of company researchers, who are able to spot within the epistemic community where the lacking skills and instrumentation reside. This is attained by attending conferences, meetings and events and keeping personal contacts with former colleagues at university[2] or at previous work experience. This "self-coordination among experts" across organizational boundaries, as

[2] According to Liebeskind et al. (1996, p.433) collaborative research representing a continuation of research programs between firm scientists and their colleagues who had remained at universities is evidently valuable to the firms, because it was the reason for their employing these scientists in the first place.

nicely put by Liebeskind et al. (1996, p.440), is allowed by employing firms since they are functional to organizational objectives, even at the cost of some misappropriation of knowledge. Oliver and Liebeskind (1998) also bring evidence of how DBFs themselves foster individual level ties between employed scientists and university scientists in order to source critical knowledge inputs.

An epistemic community (Knorr Cetina, 1981) is a collectivity of inquirers possessing the same basic knowledge - molecular biology and genetics in our case - (Cohen and Levinthal, 1990, Lane and Lubatkin, 1998) and thus sharing language, concepts, norms of inquiry, models, categories, scholarly practices (Miller and Fox, 2001, Gittelman, 2007). Personal acquaintances also enhance mutual trust[3]. Within the basic knowledge space characterizing an epistemic community, the specialised pieces of knowledge of its individual members may easily flow and be effectively exchanged. Understandability is warranted. Put simply, researchers belonging to various organizations, even though specialised in different fields, can easily understand each other, since they share the basic knowledge characterising the epistemic community, and thus are able to combine and integrate their different specialist competences.

In contrast, expanding capabilities through organizational learning requires allocating resources and formal managerial decisions to do so. It does not suffice accessing the missing skills, but a process must be undertaken in order to internalize them and transforming them into new capabilities. This needs investing in staff development, new instrumentation and facilities.

The importance of the distinction between *gaining access* to a partner's skills (e.g. relying on a partner's employees for some critical operation) and *internalizing* them as

[3] Powell et al. (1996) define the collectivity of individuals among whom exchanges of knowledge take place as a social network, stressing the importance of shared social norms of trustworthy behaviour as support to the exchanges. Preferring to stress the cognitive aspect, we refer to epistemic communities.

motivational drivers of research collaborations was first indicated by Hamel (1991, p. 84). Even in formal alliances he found that collaboration could imply “a stable division of roles based on the unique skill endowments of each partner, rather than a potentially low-cost route to replicating partner skills and erasing initial dependencies” (1991, p. 92). Later Mowery et al. (1996) clearly stated that firms, rather than using alliances to acquire capabilities, might use them to gain access to those of partnering firms, without aiming at internalizing or acquiring them. They argued that if collaborations led to acquire the partner’s technological capabilities, then over time alliances would produce increased similarity in their knowledge bases. But if they led only to access partner’s competences, partners would maintain, and possibly increase, their knowledge specialization. By examining the citation patterns in a firm's patent portfolio, these Authors could observe the changes in the relationship of one firm's technology portfolio to that of a partner firm occurring before and after alliances. By examining 792 alliances established in the years 1985-876 and including at least one US firm, they found that the capabilities of partner firms became more divergent in a substantial subset of alliances, which is consistent with the ‘knowledge accessing’ view of collaborations.

Developing this idea, Grant and Baden-Fuller (2004) contended that ‘knowledge accessing provides the predominant motive for alliance formation, especially within the knowledge-based sectors where alliance activity has been especially prevalent (e.g. pharmaceuticals, semiconductors, aerospace, telecommunications, and consumer electronics). To make clear the difference between learning and knowledge accessing, they make the example of Daimler-Benz’s initial collaboration with Swatch in designing its “Smart car”, that was motivated not by Mercedes’ desire to acquire Swatch’s precision engineering and microdesign capabilities and Swatch’s desire to acquire Daimler’s automotive know-how, but by both parties’ desire to create value through combining their separate knowledge bases. In addition, Grant and Baden-Fuller argue that the trend since the early 1980s towards the emergence of increasingly focused companies ‘is not obviously consistent with the firms using alliances to continually broaden their knowledge bases as they acquire their partners’ knowledge’. Moreover, the two approaches (respectively the *learning view* and the *knowledge accessing view* of collaborations) have differing predictions about a firm’s potential for managing multiple alliances. If alliances are about acquiring knowledge, then each firm’s number of alliances will be limited by its absorptive capacity. Conversely, knowledge accessing alliances would permit a firm to engage in extensive networks of alliances.

Similarly, Nielsen and Nielsen (2009) underline that alliance partners might intend to create value through combining their separate knowledge bases, without seeking to learn from each other, but rather willing to maintain their distinctive competences. If innovation is enabled by combining different knowledge specializations, we suggest that collaborations between the same specialized knowledge carriers should tend to be repeated any time a combination of their types of knowledge is required.

In short, according to the *knowledge accessing view* the main purpose of collaborations is likely to be overcoming the cognitive limits set by specialization and division of (intellectual) labour. By teaming with actors commanding different specializations, firms avoid spending efforts to acquire new capabilities and enlarging their cognitive scope (i.e., *they avoid* the costs of learning). They only need to be able to integrate the contributions of the various actors.

This stream of literature neither refers explicitly to the concept of CD nor makes any suggestion regarding the more likely degree of CD between partners willing to access each other’s knowledge. We believe that this connection is important, since it opens the way to a novel interpretation of the degrees of cognitive distance between partners.

In fact, if the motivation that drives interactions is to access knowledge not internally mastered, the knowledge profiles of collaborating organizations should not match and high levels of technological CD between collaborators is predictable. In contrast, according to the learning view, the likelihood of alliance formation is highest for firms that have medium levels of technological CD, that is somewhat similar knowledge profiles (see Wuyts et al., 2005, p.281). Thus we have two different predictions and interpretations of the degrees of CD between interacting organizations, which are related to the different motives underpinning diverse patterns of collaborations.

3. A methodology to measure cognitive distance

Both Wuyts et al. (2005) and Gilsing et al. (2008) use the technological classes of patents to indicate the specialization of firms and measure CD on the basis of CRTA, which is the Pearson correlation index of the distribution across technological classes of the revealed technological advantages (RTA) of each firm relative to the collaborating firms. The RTA of a firm in a particular technological field is given by the firm's share in that field, relative to its overall share of patents granted to all collaborating companies. Positive values of CRTA indicate similarity of the pattern of technological specialization of firms. Gilsing et al. (2008) apply this method to measure CD between the 116 largest companies in three industries (chemical, automotive and pharmaceutical industries) that were also establishing technology based strategic alliances and consider 400 technological classes at two-digit level. They find that their hypothesis of an inverted U-shaped effect of CD on innovation performance of firms is confirmed. Wuyts et al. (2005) apply the same index to 67 among the largest firms operating in ICT industries, considering 31 technological classes related to ICT, but do not find any relation between CD and the likelihood of alliance formation.

The first observation we make is that the correlation of the distribution across technological classes of RTA of two big size firms tells nothing about the specialization of the subunits that directly collaborate. But it is at this more micro level that the purported learning or knowledge access should take place, an insufficient absorptive capability would be an impediment to collaboration and the novelty value could be assessed. However, since this methodology cannot be applied to organizational parts (or small firms) specialized in one or, at most, few fields, a new methodology is required that allows to compare the specialization of the sub-units that are directly involved in an alliance. The second (and related) observation is that the knowledge space must be subdivided into fields of knowledge sufficiently disaggregated to permit to identify the specialized competences that characterize interacting entities.

Since formally the problem of measuring CD between firms (or other entities, like public research organizations or university departments) is identical to that faced by ecologists attempting to measure the distance of different species, once we substitute areas of competences for biological species traits, we decided to resort to measures developed by ecologists (Pielou, 1984). They have built a number of indexes of similarity or of their converse, dissimilarity or distance. While some of them can only be used for continuous or for highly variable data, others take into account only the presence or absence of certain characteristics. Furthermore, some measures are inappropriate for cases in which most of the data points are zero, as happens when one analyses small specialized firms or subunits of big organizations that typically possess only one or few competences (characteristics), among many potential ones.

We chose the index called Percentage Remoteness (PR), which is the complement of Ruzicka's similarity index (RI). According to Pielou (1984) this measure has the advantages of (i) being usable for presence, absence data and (ii) not being adversely affected by the presence of few ones and many zeros in the data. The PR measure is calculated by first computing Ruzicka's similarity index and then its complement to 100. To calculate Ruzicka's similarity index we need to compute the minimum (MIN) and maximum (MAX) for each component of the technology vectors representing the knowledge bases of the collaborating partners (Fig. 1 and equations 1 and 2).

	KB_1	KB_2	MIN	MAX
T_1	0	1	0	1
T_2	1	0	0	1
T_3	0	0	0	0
T_4	0	0	0	0
T_5	1	1	1	1
			ΣMIN = 1	ΣMAX = 3

Fig. 1. Example of steps in the calculation of Ružička's similarity index (RI) and of percentage remoteness (PR)

In the examples of fig. 1, the technology vectors representing the knowledge bases of two firms, KB_1 and KB_2, contain five component technologies (T_1-T_5). In the KB vectors the number one indicates the presence of a technology in the KB of the firm and zero its absence.

Equation (1) is Ružička's index of similarity RI.

$$RI = 100 \times \frac{\sum_{i=1}^{s} \min(x_{i1}, x_{i2})}{\sum_{i=1}^{s} \max(x_{i1}, x_{i2})} \qquad (1)$$

Equation (2) is the calculation of PR, percentage remoteness.

$PR = 100 - RI$ (2)

In this paper we shall use PR to measure cognitive distance among pairs of collaborating entities. Obviously, in order to apply it, we need to know the vector of the fields of competence that characterize each interacting actor.

Clearly, using this methodology we do not assess the extent to which the fields of knowledge possessed by the actors are similar or different, but rather we identify the domains of specialization of the various agents and count the number of fields they have in common, each field being a trait characterizing the agents. The higher the number of fields they share, the more similar they are. Put differently, what we measure is the degree of cognitive overlap of the interacting organizations or, if they are big diversified entities, of the sub-units that directly cooperate.

Finally, it is worth noting that since it is between individuals that the exchanges of knowledge ultimately take place (Grant, 1996), in theory one should consider the fields of specializations of the border-crossing researchers (and not of the organizations to which they belong) and measure cognitive distance at the individual level. Unfortunately, this degree of detail is unattainable other than in case studies, since interacting researchers are usually much more numerous than the organizations to which they belong (e.g. two firms interact but the persons involved are ten). However, if the interacting organizations are small (as is the case for most of the focal firms we examine) and possess only one or few competences, the implicit hypothesis that the fields of competences of individual researchers mirror to a great extent the organizational ones is justified[4].

[4] In order to investigate whether this assumption is acceptable we examined the correspondence of the fields of specialization of the individual researchers belonging to about 10% of the organizations collaborating with the focal firms and found that in no case the individual researchers' fields of specialization were not comprised in the employing organization list of competences.

4. Research design and data

To study cognitive distance in research collaborations, we develop an exploratory analysis of 1.244 collaborations realized by a sample of 31 Italian biotech firms with other organizations during the period 1992-2008. The data we use to trace collaborations are the scientific articles published in the journals covered by Science Citation Index Expanded (Thomson Reuters database) and co-authored by these firms and the other entities. To repeat, even though what we find is only a part of the overall innovation network involving biotech firms, joint publications reveal a substantial part of the network directly linked with the R&D activity.

Identifying the main cognitive specializations or fields of competence, of the various actors has been particularly demanding, since it required a thorough examination and comparison of various data sources. More precisely, we examined the Directory of Italian biotech firms (www.italianbiotech.com), the web sites of the firms, press articles, various internal documents, patent descriptions and the content of scientific articles. After excluding many cases for insufficient information, a sample of 31 firms which published at least one article was left. These 31 firms belong for the most part to the red biotech sector, with only 3 specialised in the green area; they are mainly independent small or medium sized companies, but 5 are affiliates of multinational companies; all of them have at least one patent registered at USPTO or WIPO.

With regard to collaborators, many of them are big, such as hospitals, universities and multinational companies. However, only a small part of these organizations is involved in the collaborations (Rosenkopf and Nerkar, 2001**)**. Starting from the name of individual co-authors, we were able to find the relevant subunits and to identify their specific knowledge bases for 845 out of a total of 1155 collaborators. 310 cases could not be examined because of absence of a website and/or public data, lack of information on knowledge bases, or because the particular unit/department involved in the collaboration was not specified.

Overall, we distinguished 25 fields of competence (see their list in the appendix), so that, for each of our firms and their collaborators we could construct a vector, constituted by these 25 components, where the presence of a field of competence is denoted by a one and its absence by a zero. Given the small size and the high level of specialization of most of the firms or of their collaborators (considering only the relevant subunits of big entities), the fields of competence vectors contained few ones and many zeros. The nature of the data we used is important, because it constrained the measure of distance we could use.

5. The structure of collaborations

F published 511 articles with 845 collaborating institutions (table I). Since various firms have the same collaborators, the number of relationships (links) realised by F (898) is higher than the total number of partners. The number of collaborations, where a collaboration is a co-publication of an F firm with any co-author, is even higher (1.244), since some relations are repeated (1,4 times on average).

The collaborating institutions (C from now on) are different kinds of organisations worldwide (34 countries in total): universities, hospitals, research institutions (including science parks, non-profit organizations, government laboratories) and firms, mainly of big or medium size, to the subunits of which we refer in the paper (as said above).

The relative importance of the above mentioned organizations as co-publishing partners of F does not vary significantly if we consider their number, or rather the number of relationships or of collaborations (table II and fig. 2).

Table I. Main data on the network of collaborations

Total # of articles published by F	511
Total # of collaborators C (# of nodes)	845
Total # of relationships (# of links)	898
Total # of collaborations (# of links*value of each link)	1.244
# of articles per firm : Average	16,3
Median	8
Modal value	2
Min. value	1
Max. value	59
# of collaborators (C) per firm : Average	29
Median	18
Modal value	3
Min. value	1
Max. value	148
Value of links : Average	1,4
Median	2
Modal value	2
Min. value	2
Max. value	31

Table II. Weight of the various types of institutions collaborating with the focal firms

Types of collaborators	**# collaborators**		**# relationships**		**# collaborations**	
	N.	%	N.	%	N.	%
Firms	92	10,9%	95	10,6%	126	10,1%
Research Institutes	97	11,5%	107	11,9%	179	14,4%
Hospitals	281	33,3%	295	32,9%	396	31,8%
Universities	375	44,4%	401	44,7%	543	43,6%
Total	**845**	100,0%	**898**	100,0%	**1.244**	100,0%

In general, universities are always the most important partner (with a share of about 44%), followed by hospitals (about 33%). The weight of firms and research organization is much lower and rather similar, ranging between 10% and 14%. These results are consistent both with those of Liebeskind et al (1996), who found that 86% of all institutions co-authoring the papers of the two firms examined were universities and other non-profit research institutions, and those of Gittelman and Kogut (2003), who uncovered that 90% of the 1.899 organizations collaborating with sample firms were universities, hospitals and research institutes, firm-firm collaborations making up a very small portion of the data. In general, the importance for DBFs of being connected with the scientific community has being widely recognized by the literature (Hicks, 1995; Cockburn and Henderson, 1998) and it is natural that if firms decide to publish, many of their papers will be co-authored with scientists of the public sector (Tijssen, 2004).

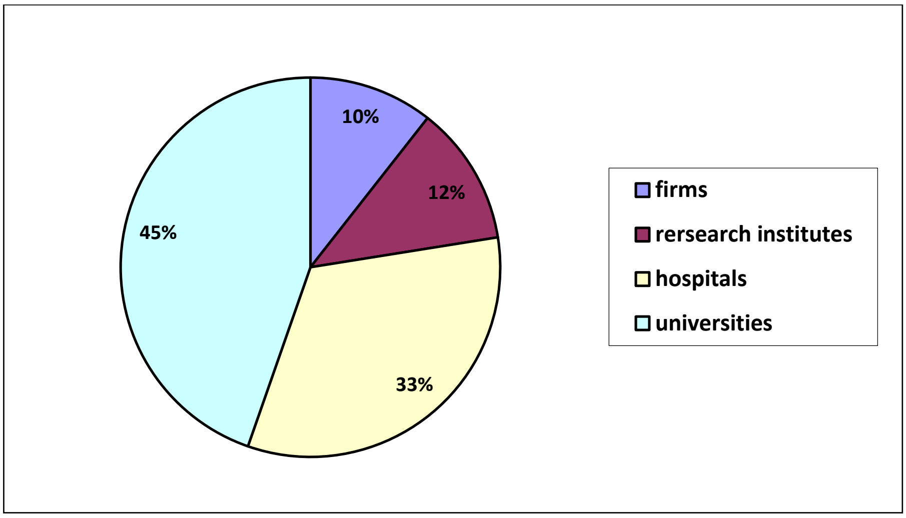

Fig. 2. Weight of the various types of collaborating institutions (shares of relationships)

Interestingly, the number of collaborators per firm is quite high (the median value is 18). For small firms such an extensive network of alliances is consistent with the knowledge accessing view (Grant and Baden-Fuller, 2004), especially considering that the degree of knowledge overlapping with the various collaborators is on average very low (as will be illustrated in the next section). The fact that small firms seek to learn so many new disciplines does not seem to make much sense.

Overall, only 39,7% of collaborations are repeated (fig. 3), but within this group the share of “strong ties”, that is of collaborations repeated more than 12 times, is significant

(20% of repeated collaborations, but only 8% of the total). This might suggest that specialised actors prefer to access repeatedly complementary knowledge carriers, in order to solve problems that need the contribution of lacking skills, rather than invest to develop those skills and to become independent.

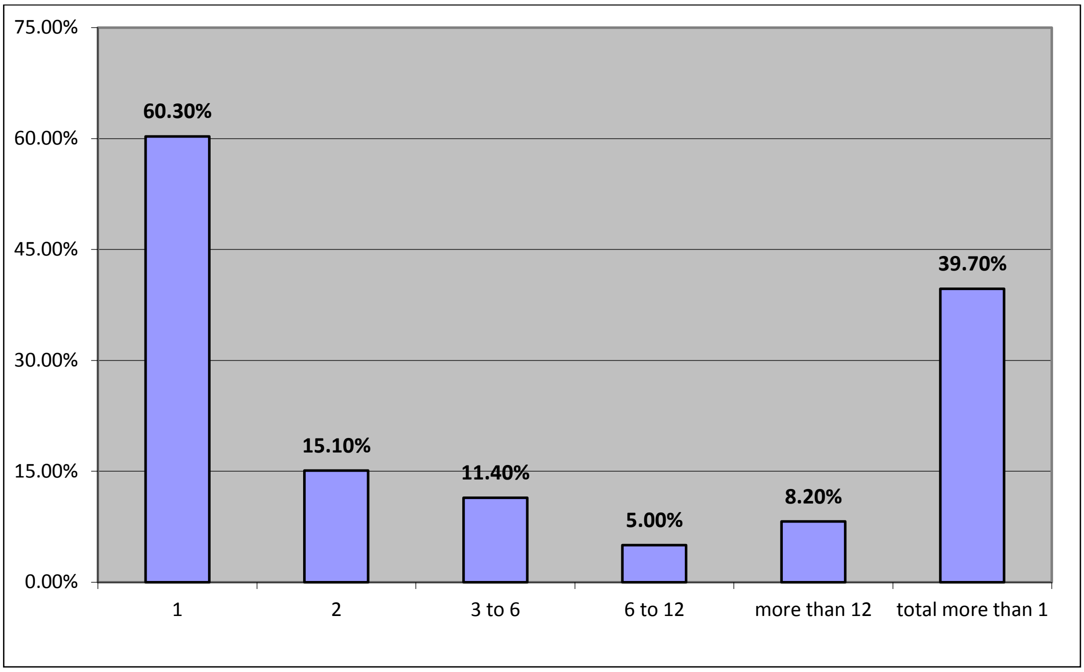

Fig. 3. Distribution of 1244 collaborations according to the number of times they are repeated

With regard to the geographic distribution of collaborations 76% are with partners located within Italy (table III). In more detail, 32,3% of collaborations are established within the same Italian region and 6,3% within the same macro area, while 37,5% with partners located in the rest of Italy. Outside Italy, European partners have a slightly higher share than partners located in the rest of the world (ROW). Within ROW, 77% of collaborations occur with partners located in the United States. It is worth noting that Gertler and Levitte (2005), in their comprehensive review of the literature on the geography of knowledge interactions in biotechnology, report a growing appreciation that also non-local (national and global) linkages are essential to successful innovation, even taking for granted the importance of proximity. Moreover, Gittelman (2007) suggests that small biotech firms are able to tap knowledge at distance, since they employ individuals belonging to epistemic communities stretching well beyond local regions.

Table III. Distribution of collaborations by geographic distance and number of times they are repeated

Geographic distance	**Distribution of collaborations by the number of times they are repeated**									
	1		2 to 6 times		> than 6 times		Total relationships (links)		Total collaborations (links* value of each link)	
	N.	%	N.	%	N.	%	N.	%	N.	%
A) Same italian region	191	25,5	42	30,9	8	66,7	241	26,8	401	32,2
B) Same Italian macroarea, but outside the region	43	5,7	12	8,8	0	0,0	55	6,1	78	6,3
A+B=C	**234**	**31,2**	**54**	**39,7**	**8**	**66,7**	**296**	**33,0**	**479**	**38,5**
Rest of Italy	279	37,2	57	41,9	4	33,3	340	37,9	466	37,5
Total Italy	**513**	**68,4**	**111**	**81,6**	**12**	**100**	**636**	**70,8**	**945**	**76,0**
Europe	122	16,3	17	12,5	0	0	139	15,5	163	13,1
Rest of the World	115	15,3	8	5,9	0	0	123	13,7	136	10,9
Total	**750**	**100%**	**136**	**100%**	**12**	**100%**	**898**	**100%**	**1.244**	**100%**

Thus, while regional embeddedness does not limit the search for a research partner, still the fact that collaborations with Italian partners account for three forth of the total seems to indicate that geographical distance and cultural proximity are important. This appears much more clearly when repeated partnerships are considered, since the average frequency of collaboration rises when the geographic distance of the partners falls.

The existence of partnerships with entities located in USA, Japan, Canada or Australia suggests that another crucial factor inducing collaborations is likely to be the distance with respect to the technological frontier of the time. This is consistent with the argument put forth by Moodysson and Jonsson (2007), namely that the convenience of local collaborations can never replace “the extreme requirements of cutting-edge specialized knowledge”, that force DBFs to seek partners in the global arena. In fact in many fields of biotechnology and medical research the frontier is located in the USA (Dosi et al., 2006) with other important organizations being located in Canada or Australia. Thus Italian biotechnology firms will opt for local knowledge whenever that is available, but will go anywhere to obtain knowledge which is scarce or unavailable locally. Of course local and international collaboration are not equivalent. The local ones may be aimed at solving recurrent problems which need continuous consultation, as shown by the very high contribution of local partnerships to repeated co-publications. On the other hand, the more expensive collaborations with a very distant partner will be used to acquire very scarce but very important knowledge. Similarly, the attractiveness of particular 'catalyst' institutions (Aygodan and Lyon’s, 2004) could also explain the collaborations with Italian universities and research institutes located outside the same region or macro area. Summarizing, direct and continuous interactions are easier in geographic proximity, but when locally unavailable knowledge becomes crucial, it does not matter how far the partner is located. Thus, even though geographic and cultural distance are likely to be barriers to collaboration, representing a “cost”, they can be compensated by the benefits arising from collaborating with particularly interesting partners.

5. Cognitive distance of collaborating partners: main results and discussion

In the great majority of cases, firms and their co-publishing partners have a very high cognitive distance (CD). The mean CD is 79,4, measured on a scale 0-100, while both the mode and the median are 100, meaning that co-publishing partners do not share any specialised cognitive field (fig.4)[5].

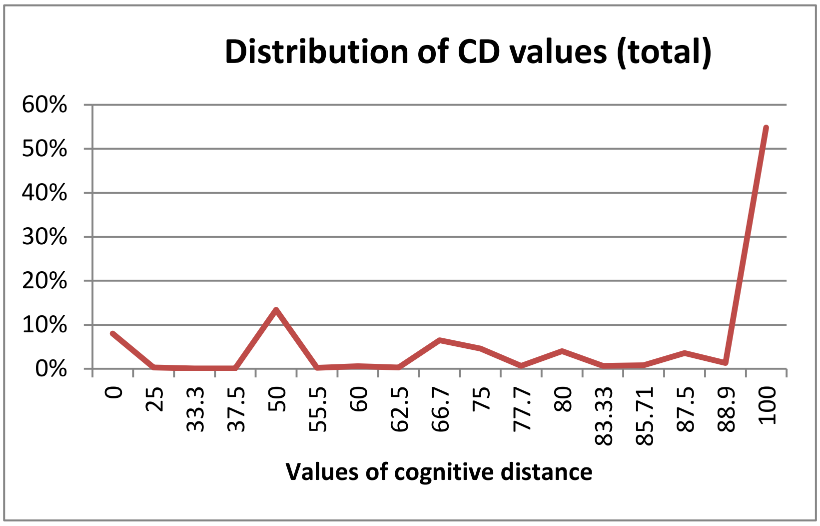

Fig. 4. Distribution of values of cognitive distance (898 cases)

These values are consistent with *the knowledge accessing view*, which predicts that the likelihood of collaborations is highest between actors with high levels of technological cognitive distance, but not with *the learning view*, that predicts medium levels of technological cognitive distance.

Clearly our results could have been affected by (i) the method used to measure CD, (ii) the way in which the cognitive fields constituting the knowledge bases of co-publishing partners are classified, (iii) the fact that the expected CD for research collaborations is not necessarily the same as for other types of collaboration, which may have also different purposes.

Although we cannot be certain that the CDs we measure are the 'true' ones, we can still expect that the high values we generally find are not an artefact of our method: a simple

[5] For a sample of 80 collaborations, we calculated CD at the level of co-authoring individuals, finding a value very similar to that calculated at the level of organizations (difference statistically not significant).

visual inspection of the data matrix displaying the competences possessed by all the co-publishing partners shows that, in the vast majority of cases, they don't have any competence in common. Thus, we consider the result obtained a realistic representation of the studied sample. Moreover, it makes sense that collaborations that yield papers are done with partners endowed with complementary specializations, and that each party is not driven by the purpose of learning each other's competence, but rather by the usefulness of accessing it.

With regard to the system used to classify fields of competence, any classification system is by definition hierarchical, since within each field of competence we can usually identify several subfields at a lower level of aggregation. Cognitive distances and costs of communicating specific knowledge depend strongly on the level of aggregation used. Real cognitive distances within a group of technological fields at a given level of aggregation (intra-group distances) should be generally smaller than the distances between two groups of technological fields at a higher level of aggregation (inter-group distances). For example, if two potential partners having competencies in biotechnology and in electronics attempt to collaborate they are likely to face much higher barriers than two partners having competencies in two different classes of biotechnology. In the second case researchers belong to the same epistemic community, so that they have the common knowledge that makes communication and integration of competences easy, while in the first case they do not. We can observe that all competencies included in our sample are medical ones, except one "green" competence, sharing a non negligible part of concepts and theories. Furthermore, most of the co-publishing firms in our sample are highly specialized and their KB contains a very small number of competencies. Even in the case of large or very large co-publishing organizations - such as universities or hospitals - the collaboration occurs with a very small subset of the organization (department, laboratory, unit, etc.) having very specialized competencies. Thus, the very high cognitive distances we observe depend on the relatively low level of aggregation we have used. Our co-publishing partners can share a lot of knowledge even if their specialization is different.

We could say that, the lower the level of aggregation at which we measure cognitive distance, the more *local* this measure is, in the sense that it indicates the relative values of the cognitive distances *across* a group of fields of knowledge at a low level of aggregation. If we wanted to find an absolute measure of cognitive distance encompassing all levels of aggregation, we would need to calibrate it with respect to the maximum possible cognitive distance between any pair of cognitive fields or subsets of knowledge. Such a measure is for the moment impossible to carry out. The local measure of cognitive distance we propose is still useful since many research collaborations occur by combination of different but not too different fields of specialization.

The third factor potentially affecting the CD values is the type of collaboration. Co-publications are but one codified result of searching and problem solving previous to the realization of a project involving some marketable outcome - a by-product of a preliminary exploratory phase - and we can expect the average cognitive distance involved in co-publishing to be different than for the joint creation of a new drug or a new plant variety. While testing such hypothesis is beyond the scope of this paper, this is an interesting topic for future research.

In summary, the dominance of large CDs in our sample of co-publications is likely to reflect the high degree of specialization of co-publishing partners and the high degree of 'local' differentiation of their knowledge, which is compatible with a very large extent of shared knowledge which allows them to communicate across the cognitive distance observed.

Coming back to the analysis of the data, we also observe that CDs vary only very slightly with the frequency of co-publications (fig. 5, fig. 6).

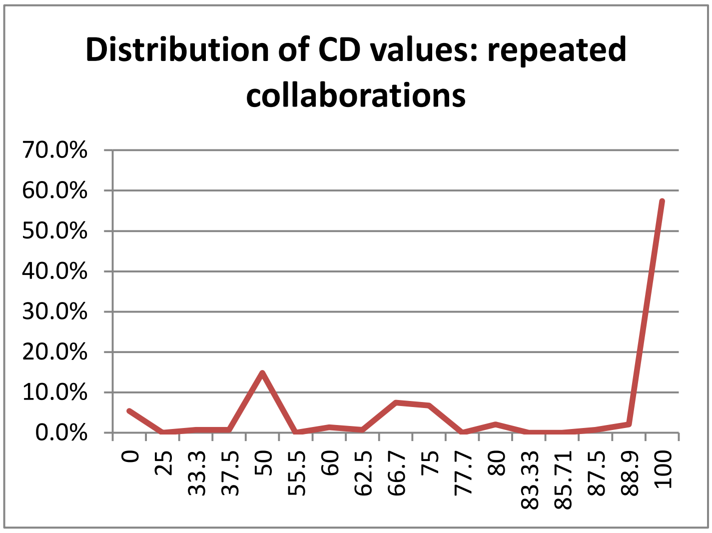

Fig. 5. Distribution of CDs values in repeated collaborations (148 cases)

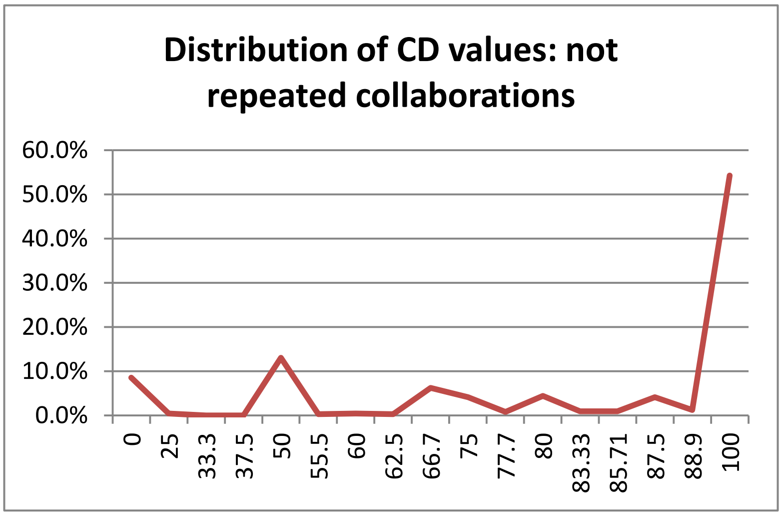

Fig. 6. Distribution of CDs values in not repeated collaborations

Only by type of partner (table. IV) does one find some difference of observed cognitive distances. The average CDs is highest for universities, followed by hospitals and it is lowest for firms and research institutes. But the differences are too small to suggest some meaning.

Table IV. Average cognitive distance between focal firms and different kinds of collaborators

	Average CD
Universities	81,37
Hospitals	79,93
Research Insititutes	74, 84
Firms	74,36
Not repeated collaborations	79,14
Repeated collaborations	80,61

Notes: The only differences statistically significant are those between universities and research institutes, and universities and firms (t-test: null hypothesis of equality of means rejected at 5%). The mode and the median values are always 100.

A final point we want to make regards the relationship between cognitive and geographic proximity. Boschma (2005) suggests that the communication difficulties due to distance in space could be surmounted by proximate cognitive capabilities between collaborating actors. More explicitly, Freel (2003) states that cognitive and spatial proximity have an inverse relationship. According to this Author, if the requisite knowledge is cognitively distant from the firm's internal knowledge base, spatial proximity becomes important in assisting effective knowledge transfer, while in the presence of cognitively proximate knowledge spatial proximity is less important.

In contrast with this literature, we find no inverse relationship between geographic and cognitive distance (table V). This can be easily explained based on our previous discussion. In particular, Italian firms need to search partners globally in order to access the specialized knowledge they lack and is not present nearby. Therefore the types of knowledge found locally are not substitutes for those found in distant places. Even if communication at distance is more costly, there is no local, cheaper source to resort to. In conclusion, the need to reach cutting-edge knowledge, which is different in terms of specialization but understandable by actors possessing the same background knowledge, loosens also the constraint of geographical proximity.

Table V. Geographic and cognitive distance between focal firms and the various types of collaborators

Geographic location of partners	N. Partners	Cognitive distance		
		Low (0-33)	Medium (33-66)	High (66-100)
Firms				
Total Italy	63	64,71%	63,16%	48,78%
Europe	29	11,76%	21,05%	28,05%
Rest of the World	26	23,53%	15,79%	23,17%
Research institutes				
Total Italy	174	71,43%	88,29%	75,96%
Europe	24	17,86%	4,50%	15,03%
Rest of the World	14	10,71%	7,21%	9,02%
Hospitals				
Total Italy	396	93,75%	51,85%	85,80%
Europe	65	0,00%	33,33%	8,88%
Rest of the World	44	6,25%	14,81%	5,33%
Universities				
Total Italy	312	64,00%	60,81%	80,97%
Europe	43	24,00%	13,51%	8,71%
Rest of the World	54	12,00%	25,68%	10,32%

7. Conclusions

In this paper we introduced a new approach to understand the degree of cognitive distance between organizations collaborating in research. Instead of assuming mutual learning as the main motive of collaborations, we argued that knowledge accessing seems a motivation more consistent with an industrial landscape characterised by firm specialization and focalization. Moreover, we proposed a classification of fields of knowledge very disaggregated, as aggregated classifications might conceal the dissimilarity between entities specialised in different knowledge niches. Relatedly, in order to understand the logic of the exchange of knowledge, we focused only on the parts of the organizations directly involved.

Since knowledge accessing instead of assimilating entails a lower need of absorptive capacity, we predicted a level of CD among collaborators higher than expected by the theory of optimal cognitive distance. Members of the same epistemic community can understand each other quite well and successfully combine their different specialised competences, since they share the same basic knowledge.

A crucial problem we faced has been coming up with an indicator suitable to measure cognitive distance at a very disaggregated level. We borrowed a measure developed by ecologists to compare different species on the basis of the number of traits they have in common, as a percentage of the total potential number of traits. In our case the traits are the knowledge fields possessed by the interacting actors.

The results of our calculations lend credence to the knowledge accessing motive. Most co-publishing partners have a high cognitive distance, the average for the whole set being 79,4 out of a maximum of 100, and the small firms examined are engaged in an extensive network of collaborations, mainly with universities. The number and variety of collaborators are not limited by absorptive capacity (within the given knowledge space) and firms 'can reconcile the benefits of knowledge specialization with those of flexible integration.' (Grant and Baden-Fuller, 2004).

However, we acknowledge that the microcosm we investigated is rather narrow. Even though the picture it provides is consistent with well known larger trends, clearly an investigation focusing on bigger firms could provide a quite different picture. In particular, the learning motive might be overwhelming, since in order to enlarge the scope of their activities big firms might intend to assimilate the knowledge of partners. This is an issue open for future investigations. Moreover, since the measure of cognitive distances we propose in this paper is not the only possible one, other measures should be tested and compared to the one we used. Finally, the results obtained for co-publications should be compared to those obtained for different types of technological collaborations, for example those aimed at the joint creation of a new drug.

Acknowledgments

The authors acknowledge the financial support of the Italian Ministry for Education, Universities and Research (FIRB, Project RISC - RBNE039XKA: ‘Research and entrepreneurship in the knowledge-based economy: the effects on the competitiveness of Italy in the European Union’) and of EC (FP6) Project, NEST-2006-PATH-Cul, CID - Cultural and Innovation Dynamics, Contract n.: FP6 – 043345). We are grateful to Roger Coronini and Giada Palamara for their help in identifying and classifying the co-publications of our Italian biotechnology firms.

Appendix

List of knowledge fields

Cardiology, Dermatology, Diabetes, Endocrinology & Metabolism, Epathology, Gastroenterology, Hemathology, Hereditary diseases, Immunology, Nephrology, Neuroscience/neurology, Oncology, Ophtalmolic diseases, Osteo-articular diseases, Otorhinolaryngology, Pain, Respiratory diseases, Rheumatology, Skeletal muscle diseases, Urology/ gynecology , Vaccines, Veterinary science/green, Virology/infectious diseases, Diagnostic supportive applications, ICT Supportive applications.

References

Audretsch, D., Stephan, P., 1996. Company-scientists locational links: the case of biotechnology. American Economic Review 86, 641-652.

Boschma, R. A., 2005. Proximity and innovation: a critical assessment. Regional Studies 39, 61–74.

Breschi, S., Malerba, F., (Eds), 2005. Clusters, Networks and Innovation. Oxford University Press, Oxford.

Cohen, W. M., Levinthal, D. A., 1990. Absorptive capacity: a new perspective on learning and innovation. Administrative Science Quarterly 35, 128–52.

Cooke, P., 2002. Biotechnology clusters as regional, sectoral innovation systems. International Regional Science Review 25, 8-37.

Dosi, G., Llerena, P., Sylos Labini, M., 2006. The relationships between science, technologies and their industrial exploitation: An illustration through the myths and realities of the so-called European Paradox. Research Policy 35, 1450-64.

Feldman, M., 1994. Knowledge complementarity and innovation. Small Business Economics 6, 363-372.

Feldman, M., 2000. Location and innovation: the new economic geography of innovation, spillovers and agglomeration. In: Clark, G., Feldman, M. P., Gertler, M. (Eds). Oxford handbook of economic geography. Oxford University Press, Oxford, pp. 373-394.

Freel, M.S., 2003. Sectoral patterns of small firm innovation, networking and proximity. Research Policy 32, 751–770.

Gertler, M. S., Levitte , Y. M., 2005. Local nodes in global networks: the geography of knowledge flows in biotechnology innovation. Industry and Innovation 12, 487–507.

Gilsing, V., Nooteboom, B., 2005. Density and strength of ties in innovation networks: an analysis of multimedia and biotechnology. European Management Review 2, 179-197.

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, 1717–31.

Gittelman M., 2007. Does Geography matters for Science-Based Firms? Epistemic Communities and the Geography of Research and Patenting in Biotechnology. Organization Science 18, 724-741.

Gittelman, M., Kogut, B., 2003. Does Good Science Lead to Valuable Knowledge? Biotechnology Firms and the Evolutionary Logic of Citation Patterns. Management Science 49, 366-382.

Grant, R.M., Baden-Fuller, C., 2004. A Knowledge Accessing Theory of Strategic Alliances. Journal of Management Studies 41, 61-82.

Hamel, G., 1991. Competition for competence and inter-partner learning within international strategic alliances. Strategic Management Journal 12, 83-103.

Hicks, D., 1995 Published Papers, Tacit Competences and Corporate Management of the Public/Private Character of Knowledge. Industrial and Corporate Change 4, 101-424.

Knorr Cetina, K., 1981. The manufacture of knowledge. Pergamon Press, Oxford.

Liebeskind, J. P., Oliver, A. L., Zucker, L., Brewer, M., 1996. Social Networks, Learning and Flexibility: Sourcing Scientific Knowledge in New Biotechnology firms. Organization Science 7, 428-443.

March, J. G., 1991. Exploration and exploitation in organisational learning. Organization Science 2, 71-86.

Miller, H.T., Fox, C.J., 2001. Epistemic Communities. Administration and Society 32, 668-685.

Moodysson, J., Jonsson, O., 2007. Knowledge collaboration and proximity. The spatial organization of biotech innovation projects. European Urban and Regional Studies 14, 115–131.

Mowery, D.C., Oxley, J. E., Silverman, B. S., 1996. Strategic Alliances and Interfirm Knowledge Transfer. Strategic Management Journal 17, 77-91.

Muller, P., Pénin, J., 2006. Why do firms disclose knowledge and how does it matter?. Journal of Evolutionary Economics 16, 85–108.

Nesta, L., Saviotti, P.P., 2005. Coherence of the knowledge base and the firm's innovative performance: evidence from the U.S Pharmaceutical industry. Journal of Industrial Economics 53, 123-142.

Nielsen, B.B., Nielsen, S., 2009. Learning and innovation in international strategic alliances: An empirical test of the role of trust and tacitness, Journal of Management Studies 46, 1031-1056.

Nonaka, I., 1994. A dynamic theory of organizational knowledge creation. Organization Science 5, 14-37.

Nonaka, I., Takeuki, H., 1995. The Knowledge-Creating Company. Oxford University Press, Oxford.

Nooteboom, B., 2000. Learning and innovation in organizations and economies. Oxford University Press, Oxford.

Nooteboom, B., Van Haverbeke, W. , Duysters, G., Gilsing, V., van den Oord, A., 2007. Optimal cognitive distance and absorptive capacity. Research Policy 36, 1016-1034.

Oliver, A.L., Liebeskind J.P., 1998. Three Levels of Networking for Sourcing Intellectual Capital in Biotechnology, International Journal of Management and Organization 27, 76-103.

Pielou, E.C., 1984. The Interpretation of Ecological Data. Wiley, New York.

Pisano, G.P., 2007. Science Business. Harvard Business School Press, Boston.

Powell, W.W., Koput, W.K., Smith-Doerr, L., 1996. Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. Administrative Science Quarterly 41, 116-145.

Pyka, A., Saviotti, P., 2005. The evolution of R&D networking in the biotech industries. International Journal of Entrepreneur and Innovation Management 5, 49-68.

Rosenkopf, L., Nerkar, A., 2001. Beyond local search: boundary-spanning, exploration, and impact in the optical disk industry. Strategic Management Journal 22, 287–306.

Rothermael, F.T., Deeds, D.L., 2004. Exploration and Exploitation Alliances in Biotechnology: A System of New Product Development. Strategic Management Journal 25, 201-221.

Storper, M., 1997. The regional world: territorial development in a global economy. perspectives on economic change. Guilford Press, New York.

Stuart, T., Ozdemir, S., Ding, W., 2007. Vertical alliance networks: The case of university–biotechnology– pharmaceutical alliance chains. Research Policy 36, 477–498.

Wuyts, S., Colombo, M.G., Dutta, S., Nooteboom, B. 2006. Empirical tests of optimal cognitive distance. Journal of Economic Behaviour and Organisation 58, 277-302.

Tijssen, R. J. W., 2004. Is the commercialisation of scientific research affecting the production of public knowledge? Global trends in the output of corporate research articles. Research Policy 33, 709–733.