Triple Helix IX International Conference
"Silicon Valley: Global Model or Unique Anomaly?"
11-14 July 2011

Sub-theme: 9. Triple Helix indicators

Cognitive and geographic distance in research collaborations: the case of Italian biotechnology firms

Margherita Balconi[a]*, Valeria Lorenzi[b], Pier Paoli Saviotti[c], Antonella Zucchella[d]

[a*]Corresponding author : Dipartimento di Economia Politica e Metodi Quantitativi, Università di Pavia, Via Ferrata 1, I27100 Pavia, Italy; tel. +39 0382985916; fax: +39 0382422583; e-mail:balconi@unipv.it;

[b]Dipartimento di Scienze Economico Aziendali, Università di Milano Bicocca, Italy

[c]INRA GAEL, Univérsité Pierre Mendés-France, Grenoble, France

[d]Dipartimento di Ricerche Aziendali R.Argenziano, Università di Pavia, Italy

Short biography of the corresponding author:
Margherita Balconi is full professor at the University of Pavia, where she teaches Microeconomics and Economics of Innovation at the Faculty of Engineering. In the period 2006-2010 she has been member of the Scientific Commettee as by the Piemonte Regional Law 4/2006, named "Regional System of Research and Innovation". Since February 2007 she is President of the Board of Polo Tecnologico Servizi, a company participated by the University of Pavia, the Municipality, the Province and the local Chamber of Commerce with the mission of creating a science and technology park in Pavia. In the last decade she published a number of articles on academic patenting, codification of knowledge, the links between basic research and innovation, new entries in the semiconductor industry and a book on science parks in Northern Italy.

1. Introduction

The main objective of this paper is to shed light on the role played by cognitive and geographic distance in biotechnology research collaborations. To this aim we investigate the scientific papers published by a sample of 31 Italian biotechnology firms and compare the knowledge base of these firms with that of the 845 organizations to which the co-authors of the papers are affiliated.

The biotechnology firms of our sample are mostly small, privately owned and have few or no patents, which in turn have typically a single applicant. For this type of firms it is very difficult both to identify the alliances in which they are involved and the fields of knowledge they master. Interestingly, since all these firms publish much more than they patent, we could rely on their co-publications in order to discover the names of the organizations with which they collaborate in research. Obviously, the collaborations revealed by co-publications are not representative of all firms' connections. For example, the research collaboration that underpins a co-publication is unlikely to involve the same kind of commitment of resources as the joint development of a new drug. However, given the central role which the generation of new knowledge plays in knowledge intensive industries, the investigation of research collaborations is interesting per se.

After having identified the collaborators of the sample of firms, the second step of our work consisted of analysing the knowledge bases of all actors, a crucial prerequisite to determine the cognitive distance between them. This was a demanding task, which required besides examining the content of the scientific articles also investigating various other sources (internet sites, journal articles, specialised databases).

As third step, we sought to come up with a methodology suitable to classify the actors' knowledge bases and the "distance" between them. We decided to resort to a methodology elaborated by ecologists to measure the distance of species based on the number of traits they share or that they do not share. Thus we do not assess the extent to which the fields of knowledge possessed by the actors are similar or different[1]. We rather identify the domains of specialization of the various agents and count the number of fields they have in common: each field is 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 agents.

Much less complex has been the issue of measuring geographic distance and of analyzing the location pattern of the collaborators. We did not find any significant interaction between knowledge distance and geographic distance, since the latter was mainly explicable on the basis of the geographical dispersion of the supply of specialized bio-medical knowledge that Italian firms needed to access.

The paper is organised as follows. In section 2 we briefly summarize the background literature on cognitive and geographic proximity. In section 3 we describe the data used and the methodologies adopted to classify the actors' knowledge bases and to measure both the cognitive and geographic distance. Section 4 illustrates the structure of collaborations among the various types of actors involved (firms, universities, hospitals, research organizations) while section 5 shows the results we obtained with regard to cognitive and geographic distance. Section 6 draws some conclusions and proposes suggestions for future research.

[1] For a "map of science" showing the connections between the various disciplines see Leydesdorff and Rafols (2009).

2. Conceptual background

Establishing collaboration with external actors is crucial in order to overcome the limitations of highly specific knowledge bases when the creation of new knowledge is involved. The literature on networks is enormous (Malerba and Breschi, 2005) and the biotech industry is widely recognised as one of the main cases of distributed innovation, where research and development take place through collaborations among organisations belonging to different scientific and business areas (Powell and Brantley, 1992; Powell et alii 1996). 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): partnerships between universities, research centres and dedicated biotechnology firms, based mostly on the exploration and creation of knowledge; and partnerships between dedicated biotechnology firms (DBF) and large pharmaceutical companies, based primarily on the exploitation of knowledge.

Geographic proximity is not in itself a pre-condition for setting up networks and for learning by interacting (Boschma 2005), but tends to establish a fertile ground for cognitive interactions which enable the sharing of tacit knowledge, while codified knowledge may also be exchanged /accessed at distance (Nonaka, 1994; Nonaka and Takeuki, 1995, Storper, 1997, Balconi et al., 2007). The new economic geography based on spillovers and agglomeration as well as regional innovation systems have also received extremely wide attention (Feldman, 2000, Cooke, 2002, Asheim and Gertler, 2005) in the literature, but to give even a cursory account of it goes well beyond the scope of this article.

We limit ourselves to observe that only when required knowledge is available locally, being located in the same area can contribute to facilitate relations between the various actors**.** In these conditions the greater possibilities of face to face contacts help create trust, while mobility of employees and sharing of experiences and technologies lead to the creation of a common knowledge base, which enhances learning processes by the members of the cognitive community in a virtuous cycle. However, the geographic distribution of innovative capabilities in recent and highly knowledge intensive sectors is far from homogeneous. In particular the USA have a considerable superiority in biotechnology with respect to European countries and important poles of competence exist in Canada, Australia, China, India to mention just a few. In these conditions it is clear that not all forms of collaboration will be locally based. Another argument to explain the possible prevalence of distant interactions is proposed by Audretch and Stephan (1996), who found that within the United States 70% of the links betweens university scientists and biotechnology companies are non local. According to them, geographic proximity is not necessary when knowledge is exchanged through formal ties – as it is in a project that delivers scientific publications –, since face to face contacts do not occur by chance but instead are carefully planned. A further argument accounting for distant relations is that, within epistemic communities, communication and cooperation rests essentially on information and communication technologies (Brown and Duguid, 1991), so that only temporary geographic proximity may suffice (Torre 2008).

On the other hand, the extent to which proximity among knowledge bases of the actors (Brown and Duguid, 1991) – that is cognitive proximity or the converse, cognitive distance - (Nesta and Saviotti, 2005; Pyka and Saviotti, 2005) is a condition for effective partnerships, is a much less studied factor, compared to geographic proximity.

Literature argues that in the process of new knowledge creation the partners involved share the expectation that the higher the cognitive distance between them, the higher the

advantage generated by an alliance (Nooteboom, 2000). That is to say that, also considering the problems linked to cognitive lock in phenomena (Lambooy, 2003), a firm can expect to learn more from an organization having the knowledge it wants to acquire rather than from one having its same knowledge. However, also the costs and difficulties a firm faces in learning a new type of knowledge tend to rise with a growing cognitive distance. Taking into account both benefits and costs, cognitive distance is considered to have an inverted U shaped relation with innovative performance (Nooteboom et alii 2007, Gilsing et alii 2008). In other word, large cognitive distance yields high opportunities for accessing new knowledge and promises a great extent of learning, but at the same time it also has a negative effect on absorptive capacity (Cohendet and Llerena, 1997), and may reach the point to preclude the mutual understanding between the interacting actors.

Thus the literature suggests that the cognitive distance is a likely determinant of the probability of existence and of the success of alliances. A firm is likely to be interested in collaboration if the knowledge it needs is so different from the one it has that it cannot develop it alone. Collaborating with another firm having a knowledge base close to the target knowledge is likely to be an effective way of acquiring it. However, large cognitive distances (CDs) cannot be expected to be very frequent, since the cost of learning would rise with a growing CD. Actual measures of CDs are extremely rare as well as empirical tests regarding these issues. Wuyts et al. (2006) measure cognitive distance in terms of technological and organisational distance on a set of inter-organisational alliances in the pharma-biotech and ICT industries. Their analysis supports the above mentioned inverse U-shaped relation.

We argue that the trade-off between the attractiveness of the target knowledge to be acquired and the cost of acquiring it can be expected to differ depending on the objective of the collaboration. For example, we could expect CD to be higher in a collaboration with a university formed by DBFs to create new knowledge (explorative alliance), than in one with a pharmaceutical firms where the objective of the alliance is the acquisition of the complementary assets (Teece 1986, Stuart et alii 2007) needed to advance and ultimately commercialize new products (exploitative alliance). We can also expect cognitive distance to vary at constant alliance objective in the course of time. As firms learn by collaborating, their CD can be expected to fall. Furthermore, average CDs are likely to be higher in the early years of a radically new technology and to fall as the technology matures.

Thus we have two main research questions to face, namely: the extent to which geographic and cognitive proximity matter in inducing research collaborations, and (as corollary) if these two distances interact between each other.

The prerequisite to address these questions is to be able to measure distance: while this bears no difficulty over the geographical space, measuring distance between collaborating organizations across the knowledge space is a problem that the economic literature has not yet solved (at least to our knowledge) and that represents the main methodological point addressed in the paper.

3. Research design, methodology and data

The main objective of our paper, as said above, is to study the influence of cognitive and geographic distance on the formation of research collaborations involving biotechnology firms. The choice of these two variables is determined by their relevance in the processes of exploration and creation of new knowledge , according to the outlined literature survey.

Thus we develop an exploratory analysis of 1.244 research collaborations realized by a sample of 31 Italian biotech firms with other organizations during the period 1992-2008. The

data we used are the scientific publications co-authored by these firms and the other entities. Even though we are aware that this is only a part of the overall innovation network involving biotech firms, we think joint publications reveal an important part of the network, directly linked with the R&D activity of the cooperating actors.

The main methodological issue we had to address was the operationalization of the concept of cognitive distance, a problem not yet solved in the literature (Gallaud and Rherrad, 2002).

Operationalising cognitive distance clearly involves two main steps. The first is identifying the main cognitive specialization, or field of competence, of the various actors, and the second is finding a methodology to measure the distance between these fields of competence.

The first step has been particularly demanding, since it required a thorough examination and comparison of various data sources. Table 1 lists the sources we used to build our dataset.

Table 1. Sources of data and information collected

Sources	*Kind of information*
The web site www.Italianbiotech.com	List of Italian biotech firms, their general characteristics, addresses
Data collected by Cresit[2]	Various data on Italian biotech firms
Web sites of the firms	Qualitative data on knowledge bases
Patent descriptions	Information on inventions and on technologies
Science Citation Index Expanded	Information on research collaborations and scientific topics

Note that the initial database comprised 96 biotech firms: missing values, unavailability of data about collaborations or knowledge bases both of the Italian firms and of their collaborators compelled us to use only part of the information, and to limit the analysis to a final sample of 45 firms. Among these, only 31 have published at least one article. These 31 firms belong mostly to the red biotech sector, with only 3 specialised in the green area; they are small or medium sized, and all of them have at least one patent registered at USPTO or WIPTO.

Overall, we distinguished 25 fields of competence, 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 (Fig.1). Given the small size and the high level of specialization of most of the firms or of their collaborators, 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.

Since formally the problem we face 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 apply several of the measures which have been developed by ecologists (Pielou, 1984). All these measures are indexes of similarity or of their converse, dissimilarity or distance. Some of them can only be used for continuous or for highly variable data, while others can also be used for presence or absence data such as the ones we have. Furthermore, even amongst the types of measures which can be used for

[2] Cresit is a research center of Insubria University, which published during the years analysed in this paper annual reports on biotech in Italy, in collaboration with FarmaIndustria, Assobiotech (Italian national associations of pharma and biotech industries) and Blossom and Associati (Italian consultancy firm).

presence or absence data, some are inappropriate for cases in which most of the data points are zero, as in our case. Another measure of similarity often used in studies of innovation has been introduced by Jaffe (1986). This measure is based on the assumption that two technologies are similar to the extent that they can be combined with the same third technology. Unfortunately, this type of measure cannot be used due to the nature of our data.

Thus, amongst all the available distance measures we chose the one called Percentage Remoteness (PR), which is the complement of Ruzicka's similarity index (RI). According to Pielou (1984, pp. 43-44 and 55-57) 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).

Technologies	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

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

In the examples of figure 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.

$$PR = 100 - RI \qquad (2)$$

Measuring geographic proximity among the collaborating entities does not present particular methodological problems. We have operationalised it through a variable whose values are based on the closeness of collaborators to the location of the firms. More precisely, we have distinguished five different values:

- 0 if the firm (F) and the collaborating organisation (C) are located in the same region of Italy;
- 1 if F and C are located in the same macro area (North-West, North East, Centre, South, according to ISTAT classification), but in different regions;
- 2 if F and C are located in Italy, but in different macro-areas;
- 3 if C is located in Europe;
- 4 if C is located in the rest of the world.

4. The structure of collaborations

F published 511 articles with 845 collaborating institutions (table 2). 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).

Table 2. 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

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.

The 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 3 and fig. 2). 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%[3].

Table 3. Weight of the various types of institutions collaborating with the Italian biotech firms of the sample

	# Collaborators		**# Relationship**		**# 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%	**1244**	100,0%

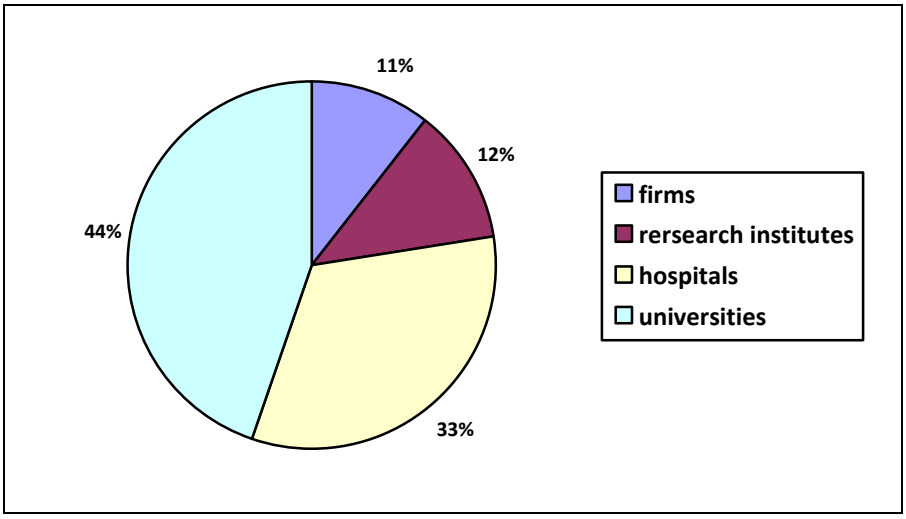

Figure 2. Weight of the various types of collaborating institutions (shares of relationships)

Overall, 39,7% of collaborations are repeated (fig. 3) and the share of "strong ties", that is of collaborations repeated more than 12 times, is relevant (20% of repeated collaborations).

[3] The most important partners are University Federico II of Naples, University Statale of Milan and University La Sapienza of Rome, CNR among research institutes and S.Raffaele of Milan among hospitals.

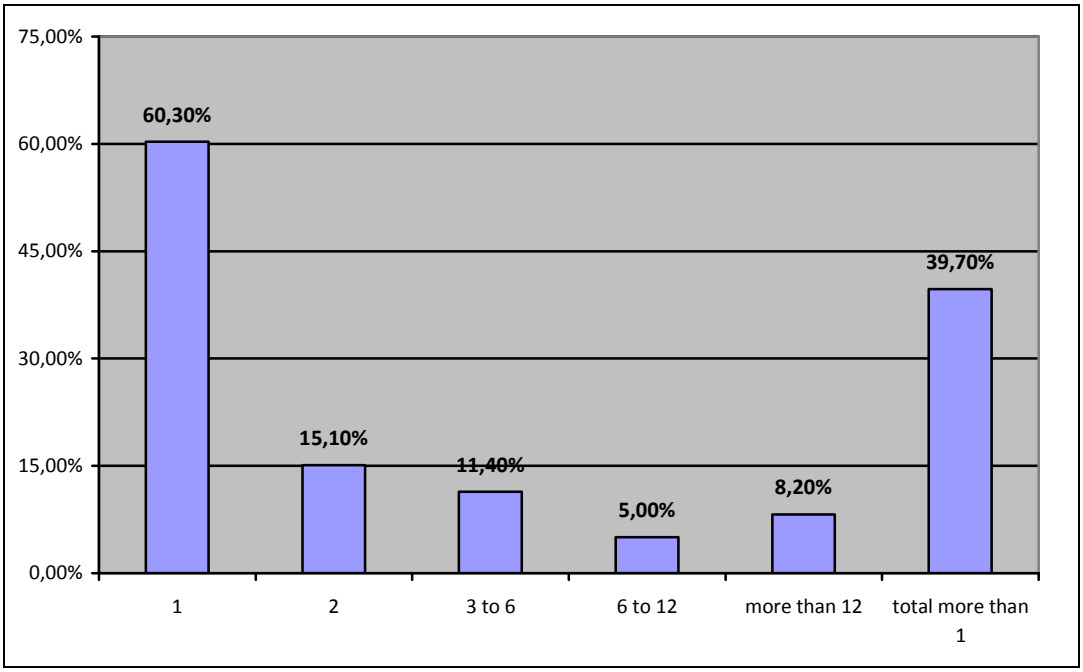

Figure 3. Distribution of collaborations according to the number of times they are repeated

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

5.1.Cognitive distance

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 the co-publishing partners do not share any specialised cognitive field (fig.4).

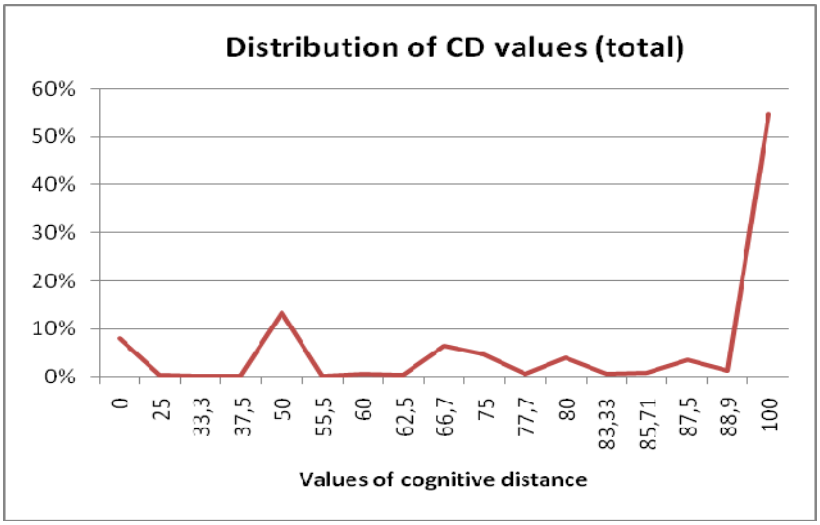

Fig.4 Distribution of values of cognitive distance (898 cases)

These values seem quite high with respect to what one should have expected from the literature on optimal cognitive distance, according to which a median value between 0 and 100 should have prevailed (Nooteboom, 2000 and 2007; Brown and Duguid, 1991, Cohendet and Llerena, 1997) and a distribution of values with the shape of an inverted U.

However, the high CD values found might not be simply considered as evidence that the advantage for each partner of accessing a different specialization greatly outweighs the costs of communication, since one needs to take into account the factors which could have affected our measurements. In fact, 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 co-publications is not necessarily the same as for other types of collaboration.

As we previously pointed out, there are several possible measures of cognitive distance. Not all of these measures are suitable for every data set. The measure we chose (PR) is suitable for a data set in which the KB vector of each collaborating partner tells us about the presence or absence of the set of specializations characterising altogether the group of collaborators studied. Concerning the data, we furthermore know that, in its previous use, this distance measurement turned out not to be adversely affected by the presence of many zeros (absence) and of few ones (presence) (Pielou, 1984). 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 visual inspection of the data matrix displaying the competences possessed by all the co-publishing partners show 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.

Another source of influence on the CD values is the system used to classify fields of competence. Any such classification system is by definition hierarchical, in the sense that it includes competences at different levels of aggregation. Within each field of competence we can usually identify several competencies at a lower level of aggregation. We can expect cognitive distances and costs of communicating specific knowledge to depend strongly on the level of aggregation used. We can also expect cognitive distances and communication costs to rise with a growing level of aggregation. To put it differently, the 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. 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, in general we can expect that the very high cognitive distances we observe occur for a relatively low level of aggregation. Our co-publishing partners can share a lot of knowledge and differ in a very limited subset of their KB. 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 technological alliances 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. The collaborations we are examining do not represent the full network of partnerships established by our sample of firms. It comprises only those relations among actors who conduct research together, and are thus co-authors of the resulting publication. Co-publications are but one codified result of learning previous to the realization of a project involving some marketable outcome - a by-product of a preliminary exploratory phase - and we can expect the shared experiences and knowledge required for a successful collaboration of this kind to be more limited than for the joint realization of a marketable outcome. In other words, we can expect the average cognitive distance involved in co-publishing to be different and possibly higher than for the joint creation of a new drug or a new plant variety. Of course, this is more an hypothesis to be tested than an accepted result which can be used as an explanation of new findings. 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-publication and with the type of partner (fig.5 – 10, tab.4)

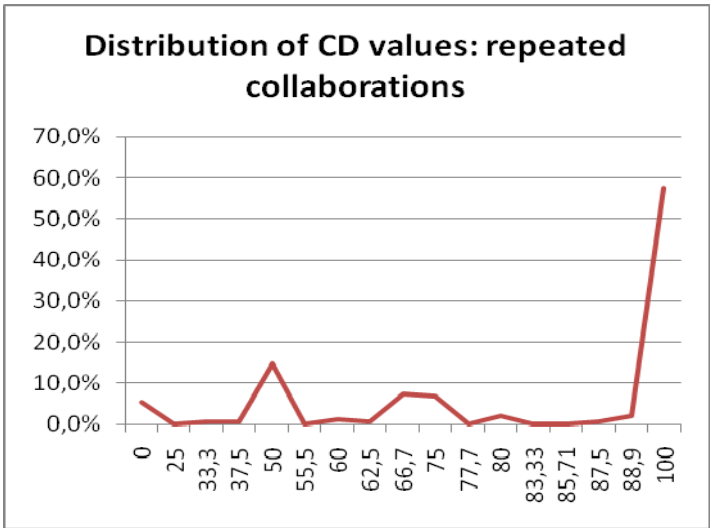

Fig.5 Distribution of CDs values in repeated collaborations (148 cases)

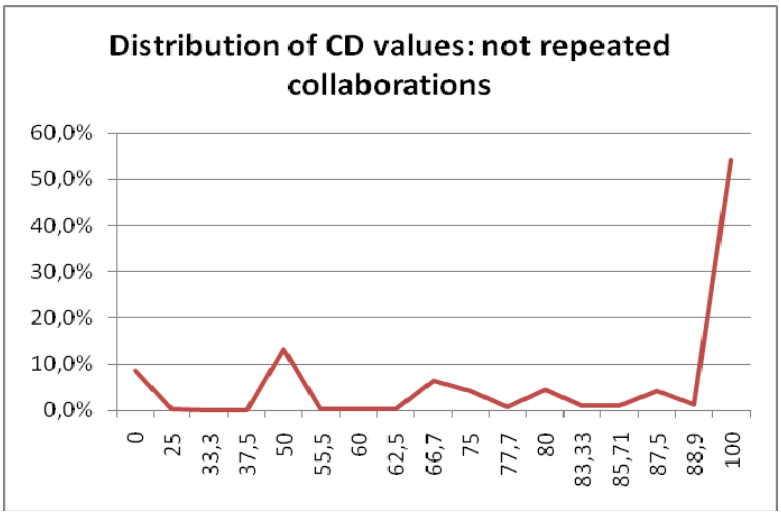

Fig.6 Distribution of CDs values in not repeated collaborations (750 cases)

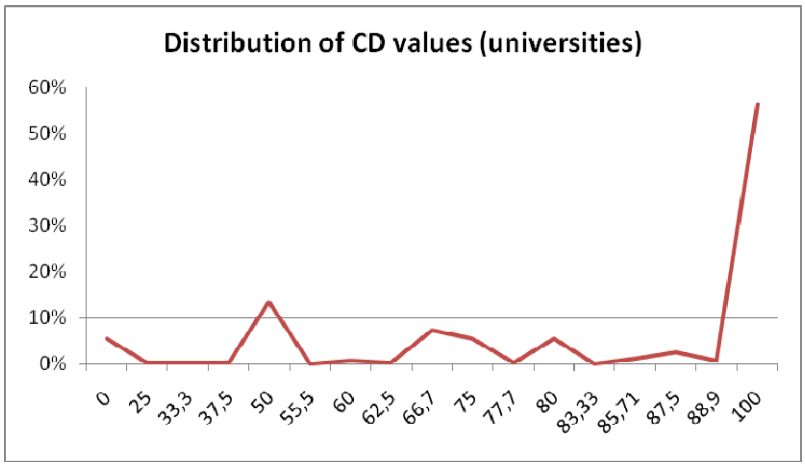

Fig.7 Distribution of CDs values in with universities (375 cases)

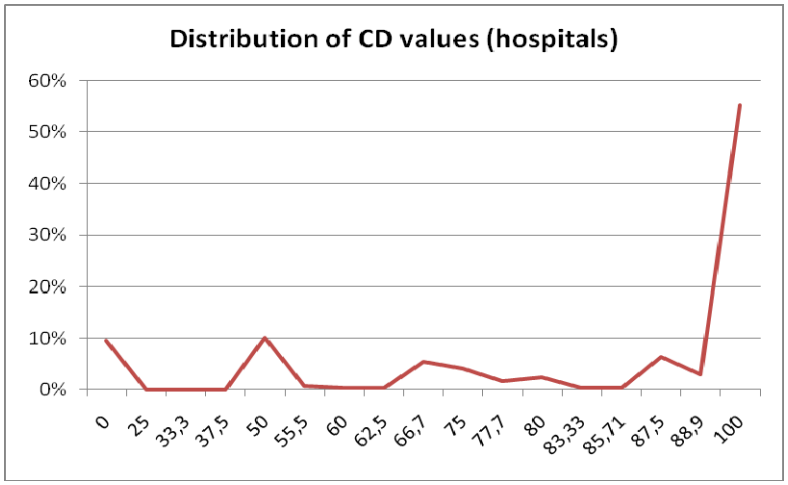

Fig.8 Distribution of CDs values in with hospitals (281 cases)

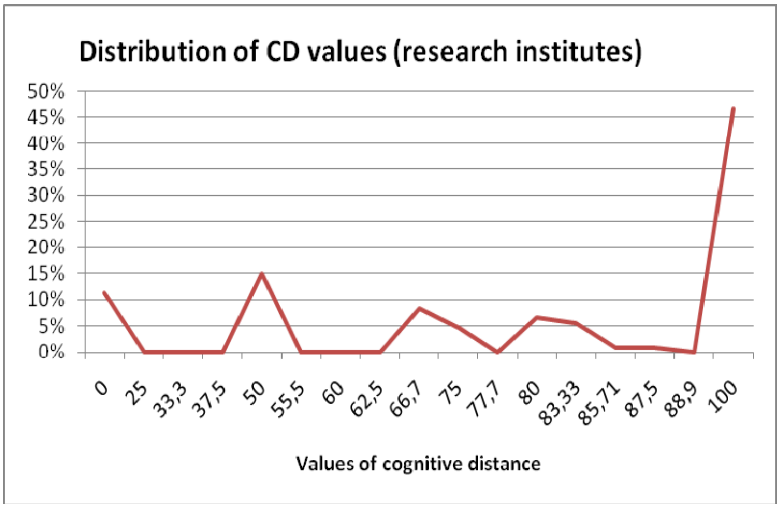

Fig.9 Distribution of CDs values in with research institutes (97 cases)

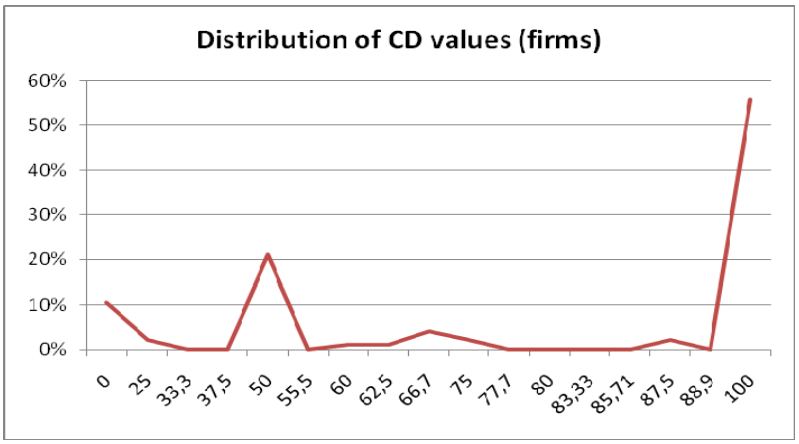

Fig.10 Distribution of CDs values in with firms (92 cases)

Tab 4. Average cognitive distance between 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

Note: The mode and the median values are always 100.

Only by type of partner 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 (table 4). Although our interpretation is rather tentative, this order seems consistent with the idea that cognitive distances tend to be higher for the more explorative types of collaborations, which should be those with universities.

5.2.Geographic distance

With regard to geographic distance, 71% of relationships and 76% of collaborations are with partners located within Italy (table 5). 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. In particular, 77% of collaborations with entities of ROW occur with partners located in the United States.

Table 5. Geographic distance

Geographic distance	**Relationship**		**Collaborations**	
	N.	%	N.	%
A) Same italian region	241	26,84%	401	32,23%
B) Same Italian macroarea, but outside the region	55	6,12%	78	6,27%
A+B=C	296	32,96%	479	38,50%
Rest of Italy	340	37,86%	466	37,46%
Total Italy	636	70,82%	945	75,96%
Europe	139	15,48%	163	13,10%
Rest of the World (ROW)	123	13,70%	136	10,93%
Total	**898**	100,00%	**1244**	100,00%

Since less than 40% of the collaborations occur in Italy within the same region or the same macro area, local innovation systems seem not to have a major influence on the formation of partnerships, at variance with the literature (Cooke 1998 and 2002; Storper 1997, Feldman 2000). Moreover, a similar percentage of collaborations occurs with Italian partners located in the rest of Italy. Thus, while it appears that 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 (table 6).

Table 6. Geographic distance and collaborations

Geographic distance	**Distribution by number of collaborations**									
	1		2		3 to 6 times		6 to 12 times		> than 12 times	
	N.	%	N.	%	N.	%	N.	%	N.	%
A) Same italian region	191	25,5%	30	31,9%	12	28,6%	5	62,5%	3	75,0%
B) Same Italian macroarea, but outside the region	43	5,7%	7	7,5%	5	11,9%	0	0,0%	0	0,0%
A+B=C	234	31,2%	37	39,4%	17	40,5%	5	62,5%	3	75,0%
Rest of Italy	279	37,2%	39	41,5%	18	42,9%	3	37,5%	1	25,0%
Total Italy	*513*	*68,4%*	*76*	*80,9%*	*35*	*83,3%*	*8*	*100,0%*	*4*	*100,0%*
Europe	122	16,3%	13	13,8%	4	9,5%	0	0,0%	0	0,0%
Rest of the World	115	15,3%	5	5,3%	3	7,1%	0	0,0%	0	0,0%
Total	***750***	*100,00%*	***94***	*100,0%*	***42***	*100,0%*	***8***	*100,0%*	***4***	*100,0%*

Notwithstanding the importance of geographic and cultural proximity, the existence of joint research 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. In biotechnology and medical research the frontier is located in the USA (Dosi, Llerena, Sylos Labini, 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. In other words, what we suggest is that 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.

Table 7. The geographic distribution of collaborations by type of partner

	Firms	**Research institutes**	**Hospitals**	**Universities**
Same Italian region	23,2%	35,5%	34,9%	19,5%
Same Italian macroarea	4,2%	9,3%	4,7%	6,7%
Rest of Italy	14,7%	23,4%	46,8%	40,6%
Italy	***42,1%***	***68,2%***	***86,4%***	***66,8%***
Europe	32,6%	18,7%	7,8%	16,2%
Rest of the world	25,3%	13,1%	5,8%	17,0%
	100,0%	100,0%	100,0%	100,0%

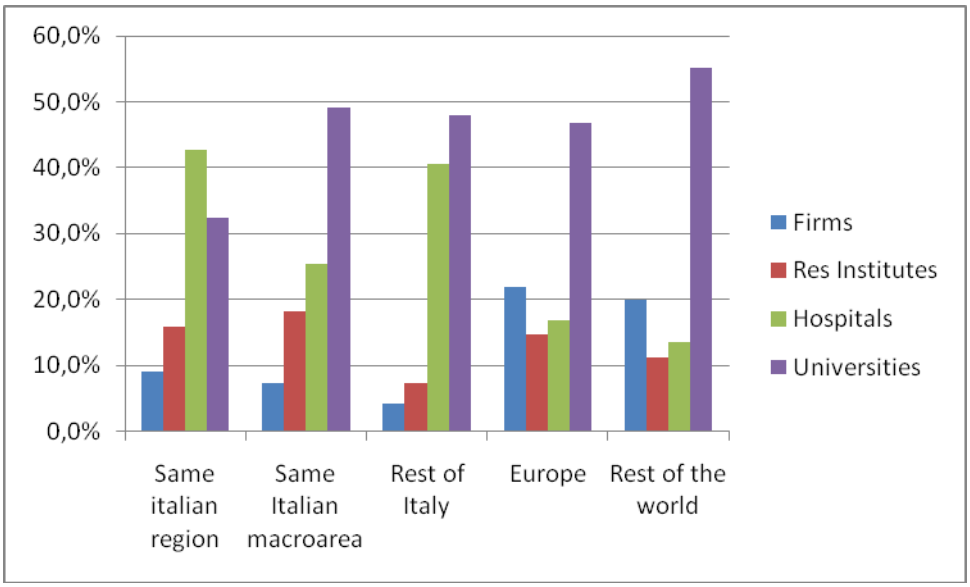

Fig.11. The importance within each area of the various collaborators.

The non equivalence of the collaborations with different types of partner is confirmed by the different distribution of co-publications in various geographical areas.

While universities are main partner in any area, it is outside of the same region that their role is particularly important, especially in the ROW (United States in particular) (table 7 and fig. 11). At a first inspection this finding might seem to contradict what we could have expected from the literature, which argues that geographic proximity with universities is fundamental in terms of localised knowledge spillovers for biotech firm (Jaffe 1989 and 1993, Vedovello, 1997). A number of factors could affect these apparently divergent results. First, we should remember that, with the exception of Molmed, spin-off of San Raffaele university, the 31 Italian biotech firms of our sample were not born as university spin-offs. Thus, to establish contacts with local rather than other Italian universities would not have been the obvious first choice for them. The important role of local research institutes can probably partly explain the relatively limited use of local universities in co-publications. Amongst the types of co-publishing partners research institutes are the closest to universities in terms of objectives and procedures. In other words, although not identical, they are reasonably close substitutes. Thus, if the local research system contains a high percentage of research institutes

it is not surprising that they account for a significant share of co-publications. One should not forget that national research systems differ considerably with respect to the relative importance of universities and research institutes: for example, research institutes are less important in the USA and in the UK than in France or Germany. Thus, the important role of local research institutes in co-publications is either due to the local organisation of research or to their intrinsic characteristics (e.g. pattern of specialisation, more applied research then universities etc).

With regard to collaborations with firms, the very high frequency of co-publications with firms located in the rest of the world and in Europe indicates that the firms of our sample, when looking for partners with an objective closer to exploitation than exploration, have to resort to foreign firms, for the simple reason that in Italy domestic big pharmaceutical firms do not exist.

At this point one has to mention the situation of biotechnology in Italy. As already pointed out, modern biotechnology was created in the USA where the frontier of knowledge is still located. During the 1970s and 1980s DBFs were an almost uniquely USA phenomenon. In Europe the number of DBFs only started rising substantially during the 1990s. In Italy the growth in number of DBFs only started in the 2000s (Blossom Associati, 2008). Thus, if the most advanced European countries in biotechnology were latecomers, Italy was a late latecomer. This is important, because the strategies required to enter an industry are likely to vary according to the period of its life-cycle. Thus, in the USA, during the 1970s and the 1980s when modern biotechnology was in its infancy, a very high percentage of DBFs were founded by scientific entrepreneurs (Zucker et al, 1998; Audtretsch and Stephan, 1996, Oliver, 2004), while this does not seem to be the case at all in Italy, where most entries started taking place in the 2000s when at least some subsets of biotechnology were already maturing.

With respect to this point we have to bear in mind that in biotechnology we can identify two generations linked to recombinant DNA and monoclonal antibodies and to genomics respectively (Saviotti, Catherine, 2008). Within the first generation, R&D collaborations had virtually disappeared by the end of the 1990s, while marketing agreements continued. If we assume that the end of R&D collaborations in the first generation implies the onset of maturation of this subset of biotechnology, it follows that only the second generation linked to genomics still represented the frontier by the beginning of the 2000s. This situation would open possibilities for different types of collaboration in the two generations. In particular we might expect a late latecomer like Italy to opt for incremental innovations in the more maturing subset of biotechnology, which is in the first generation, together with local partners and to participate in innovations near the technological frontier of the time with advanced foreign partners.

In order to test the proposition that the collaborations with non Italian, and in particular with ROW co-publishing partners are different from those with Italian partners, we constructed a list of the journals in which such co-publications appeared and of the themes describing the co-publications. We found no differences when comparing the lists of co-publications of the firms of our sample with other firms, research institutes, hospitals or universities without taking into account their geographical location. On the other hand, when we compared the distribution of the co-publications with Italian partners to those with non-Italian partners (external co-publications) by means of the key words associated with each publication we found that they differed considerably. Some examples of these differences are shown in Table 8. If we bear in mind that the percentages of each keyword - in the total of Italian and of external co-publications - represent the pattern of specialisation of each set of co-publications, we can see that the Italian and external co-publications specialise in different

subsets of knowledge. For example, “haematology” - the most common keyword associated with all Italian co-publications - accounts for 16.0 percent of all Italian co-publications but only for 3.8 percent of the external ones. On the other hand, “developmental biology” accounts for 4.0 percent of external co-publications but for 0.45 percent of the Italian ones and “behavioural sciences”, which account for 1.1 percent of external co-publications, are absent in the Italian ones. Thus, Italian biotechnology firms use co-publications with distant partners to acquire types of knowledge different from those which they can obtain by co-publishing with close by partners.

Table 8. Pattern of specialisation of the co-publications of the firms of our sample with Italian and with non-Italian partners, as detected by the keywords associated with the articles

Field of Knowledge (Keyword)	**Percentage of Italian co-publications**	**Percentage of non-Italian co-publications**
Haematology	16.0	3.8
Biochemistry and molecular biology	12.6	25.0
Neurosciences	4.8	7.7
Developmental biology	0.45	4.0
Behavioural sciences	0	1.1

Note : Only the examples showing the greatest differences are displayed in this table.

Table 9 shows no direct relationship between geographic and cognitive distance. Such lack of a direct relationship could have been expected from our previous discussion. As we pointed out, the choice of a foreigner, or, more generically, of a geographically distant partner, is likely to be dictated by the proximity of this partner to the technological frontier of the time in the desired sub-field of knowledge. The distance from the technological frontier of the time needs to be clearly distinguished from the cognitive distance that we measure. To understand the difference between the two we can imagine to represent the different sub-fields of knowledge in which we are interested on an horizontal axis ranking them in order of growing dissimilarity (or of growing cognitive distance) and of representing on a vertical axis the technological capabilities of each country or organisation in each sub-field of knowledge on a scale ranging from zero (0) to 100, where 100 would be the frontier. First, we cannot expect any direct relationship between cognitive distance (CD) and distance from the technological frontier (DF). We can only expect firms and research organisations in an imitating country to choose at least some of their partners in the country and in the organisations that are as close as possible to the technological frontier of the time in the desired sub-field of knowledge. The lack of a direct relationship between geographic and cognitive distance follows from the lack of a corresponding relationship between cognitive distance and distance from the technological frontier of the time.

Table 9. The relationship between geographic and cognitive distance for the firms of our sample and their collaborators

Geographic location of partners	# 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%

6. Summary and conclusions

In this paper we studied the influence of cognitive and geographical distance on the collaborations of a sample of Italian biotechnology firms. Amongst the co-publishing partners of our firms we distinguish universities, research institutes, hospitals and firms. Such co-publishing partners are located all over the world although their distribution is not uniform. By using various sources of data we have been able to assign to each of our firms and of their co-publishing partners some field of knowledge on the basis of which we have been able to construct a competence vector for each co-publishing partner. This vector gives information about the presence or absence of the fields of knowledge of a given range in the knowledge base of each of the firms and organisations studied. We considered a large number of measures of similarity and of distance mostly used by ecologists to measure the similarity of animal species, since their objective is identical to ours, once we substitute biological traits with fields of knowledge. Amongst the various measures of distance available we chose one called percentage remoteness (PR), because it was the most appropriate for a data set (i) containing information about the presence or absence of given technologies and (ii) in which the fields of knowledge vectors of each firm and organisation studied contain many zeros (absences) and few ones (presences). The results of our calculations show that most co-publications have a high cognitive distance, the average for the whole set being 79,4 out of a

maximum of 100. In general we can expect firms collaborating to acquire new knowledge to choose partners with a knowledge base different from theirs and closer to the target knowledge they want to acquire. However, while the advantage of the collaboration for the learning firm may be expected to increase with the cognitive distance, collaboration costs may be expected to rise in the same direction. Thus, one should expect the observed cognitive distances to reflect a trade off between advantages and costs of collaboration (Nooteboom, 2000). It may seem that the high cognitive distances we measure are not consistent with this argument. However, any measure of cognitive distance depends on the level of aggregation at which fields of knowledge are defined. When the fields of knowledge are defined at a very low level of aggregation, firms and organizations can specialise in a set of fields different from that of any of their partners while sharing with them a wide range of knowledge. In other words, any measure of cognitive distance will always be 'local' in the sense of measuring distances as a percentage of the maximum possible within a narrow range of knowledge. Absolute measures of cognitive distance could only be calculated for a set including fields of knowledge at all possible levels of aggregation. We conclude that the cognitive distances we observe are large because the organizations we study are highly specialized within a narrow range of knowledge, which allows them to collaborate with partners sharing a lot of background knowledge but having competencies different from theirs.

The cognitive distances we observe vary, although not a lot, with the type of co-publishing partner. Collaborations with partners with the highest cognitive distances are universities. Thus, high cognitive distances seem to be more frequent the more the collaborating organisations are exploration oriented. The distribution of co-publications by geographical area shows that about three forth of the co-publications are with Italian partners and almost one half with partners from the same region or macro-area within Italy. An interesting result of our analysis has been obtained by studying jointly the distributions by kind of partner and geographical distances. Universities are the most important partner in any area, but they are also the less important in the same Italian region; conversely, hospitals and research institutes are especially important in the same Italian region - or more generally in Italy - while firms are so outside Italy. These differences are likely to reflect the organization of the Italian research system, where also research institutes are an important actor in applied research. However, we interpreted these variations as arising from the different roles which can be played by collaborations with close by or with very distant partners. The geographical distribution of competencies in biotechnology is by no means uniform. For example, biotechnology firms, universities and research institutes in the USA can be expected to be much closer to the technological frontier than Italian ones. The distance of potential collaborators from the technological frontier of the time is likely to be a very important factor affecting the choice of partners by biotechnology firms. Such choice is likely to be dictated by the balance between costs and benefits of the collaboration. Collaborations with partners located in very distant geographical areas but very close to the technological frontier will in general be more expensive but impossible to replace with local collaborations, given the scarcity of the knowledge required. On the other hand, local collaborations can be used to acquire or improve knowledge required in the everyday practice of research and development. Thus, in general one cannot expect alliances with local or with distant partners to be substitutable but to play a systematically different role. To test this idea we compared the co-publications of the firms of our sample with Italian and with non Italian partners. To do this we calculated the fraction of co-publications corresponding to the keywords associated with each publication in the two groups. The distribution of co-publications by keyword, which represents the relative patterns of specialisation of the two groups, is considerably different

for Italian and non Italian collaborations. Thus, Italian biotechnology firms use alliances with Italian and non Italian partners to look for different types of knowledge.

The previous results show that in order to understand research partnerships in biotechnology (but equally in other high technology sectors) we have to take into account not only cognitive and geographical distances but also the distance from the technological frontier of the time. As a consequence the distribution of the alliances of the biotechnology firms across countries will also depend on the distance of the country with respect to the technological frontier. Firms based in a country far behind the technological frontier of the time will either need to position themselves in already maturing subsets of biotechnology or to have a fraction of their alliances in countries on or near the technological frontier.

Our study of the co-publications of a sample of Italian biotechnology firms gives some interesting results, but raises a number of issues for further investigation. First, since the measure of cognitive distances we proposed in this paper is not the only possible one, other measures should be tested and compared to the one we used. Second, 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. Furthermore, our results suggest that technological collaborations can evolve during the life cycle of the technology considered as it diffuses from the originating country to imitating countries. Thus, the mechanisms of research collaborations in biotechnology should be compared for different countries and in different periods of the life cycle of the knowledge field.

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.

References

Asheim, B., and M.S. Gertler. 2005 The Geography of Innovation: Regional Innovation Systems in: *The Oxford Handbook of Innovation*, Oxford University Press, Oxford.

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

Aydogan, N. and T. P. Lyon. 2004. Spatial proximity and complementarities in the trading of tacit knowledge. *International Journal of Industrial organisation* 22: 1115-1135.

Balconi, M., A. Pozzali, and R. Viale. 2007. The "codification debate" revisited: a conceptual framework to analyze the role of tacit knowledge in economics. *Industrial and Corporate Change* 16: 823–849.

Blossom Associati. 2008. *Biotecnologie in Italia 2008. Analisi strategica e finanziaria*, Blossom Associati, Milano.

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

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

Brown, J. S., and P. Duguid**.** 1991. Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation. *Organization Science* 2: 40–57.

Cohendet, P., and P. Llerena. 1997. Learning, technical change and public policy: how to create and exploit diversity. In *Systems of innovation. Technologies, Institutions and Organizations*, C. Edquist (ed.): 223-234, Francis Pinter, London.

Cooke, P. 1998. Introductions: origins of the concept. In *Regional Innovation Systems*, ed. H. Braczly, P. Cooke and M. Heidenreich: 2-25, UCL Press, London.

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

Dosi G., P. Llerena and M. Sylos Labini. 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: 1450-1464.

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

Gallaud, D., and Rherrad, I. 2002. Knowledge production and patterns of proximity. Paper presented at the DRUID Summer Conference on '*Industrial Dynamics of the New and Old Economy - Who is embracing whom?'*, Copenhagen/Elsinore 6-8 June 2002.

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

Gilsing, V., B. Nooteboom, W. Vanhaverbeke, G. Duysters, and A.van den Oord. 2008. Network embeddedness and the exploration of novel technologies: Technological distance, betweenness centrality and density. *Research Policy* 37: 1717–1731.

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

Jaffe, A. B. 1989. Real effects of academic research. *American Economic Review* 79: 957-970.

Jaffe, A. B., M. L. Trajtenberg and R. Henderson. 1993. Geographic localisation of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics* 63: 577-598.

Lambooy, J. G. 2003. The role of intermediate structures and regional context for the evolution of knowledge networks and structural change, *Papers on Economics and Evolution,* Max Planck Institute, no. 0309. Jena.

Leydesdorff L. and I. Rafols. 2009. A Global Map of Science Based on the ISI Subject Categories. *Journal of the American Society for Information Science and Technology*, 60: 348–362.

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

Nesta, L., and P.P. Saviotti. 2005. Coherence of the knowledge base and the firm's innovative performance: evidence from the U.S Pharmaceutical industry. *Journal of Industrial Economics* 53: 23-142.

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

Nonaka, I., and H. Takeuki. 1995. *The Knowledge-Creating Company*, New York and Oxford, UK: Oxford University Press.

Nooteboom, B. 1998. Innovation and Inter-firm Linkages: New implications for policy. *Research Policy* 28: 793-805.

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

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

Oliver A. 2004. Biotechnology entrepreneurial scientists and their collaborations. *Research Policy* 33: 583-597.

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., W.K. Koput and L. Smith-Doerr. 1996. Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology. *Administrative Science Quarterly* 41: 116-145.

Powell, W. W., and P. Brantley. 1992. Competitive cooperation in biotechnology: learning through networks? In *Networks and organisations: structure, form and Action*, ed. R. Eccles. Harvard Business School Press, Boston.

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

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

Saviotti, P.P., and Catherine D. 2008. Innovation networks in biotechnology. In *Handbook of Bioentrepreneurship,* ed. H. Patzeltand T. Brenner: 53-82. Springer Verlag, Heidelberg.

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

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

Teece, D.J., 1986. Profiting from technological innovation - implications for integration, collaboration, licensing and public policy. *Research Policy* 15: 285–305.

Torre A. 2008. On the Role Played by Temporary Geographical Proximity in Knowledge Transmission, *Regional Studies*, 42: 869–889.

Vedovello, C. 1997. Science parks and university-industry interaction: geographical proximity between the agents as a driving force, *Technovation* 17, no.9: 491-502.

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

Zucker L.G., M.R. Darby and M.B. Brewer. 1998. Intellectual human capital and the birth of US biotechnology enterprises. *American Economic Review* 88: 290-306.