Research collaboration networks in biotechnology: exploring the trade-off between institutional and geographical distance

Rosamaria D'Amore[*] Roberto Iorio[†] Sandrine Labory[‡] Agnieszka Stawinoga[§]

Provisional draft – 3 October 2012 – Please, do not quote

Abstract. This paper focuses on the way institutions involved in scientific collaboration deal with geographic and institutional distance. Geographic distance refers to the location of research collaborators in different places, institutional distance refers to the diversity of institutions whom research collaborators belong to.

There are both barriers and advantages to overcome such distances and the balance between such costs and benefits may induce institutions to consider institutional and geographic distance as substitutes; therefore a trade-off between them is possible. As both institutional and geographic distance in research collaboration may be related to the nature of the exchanged knowledge, the effect of the nature of the research (if basic or applied) on the trade-off is also considered. Besides, the relationship between the two kinds of distance is likely to depend on the frequency of collaborations, therefore the "stability" of relationships is also taken into account.

The phenomenon of collaborations among different institutions (firms, universities, hospitals and research centres) is seen through the lens of co-autorship of scientific publications in the Italian "red" biotech sector.

As empirical tools, we adopt some indexes built in the context of the social network analysis (the E-I index and the equivalence coefficient) but previously used in different applications. The social network analysis is complemented by a regression analysis.

Keywords: University-industry collaboration; Co-authorships; Spatial distance; Institutional distance; Knowledge flows.

JEL Classifications: L65; O33; O31

[*] Universisty of Salerno, Department of Economics and Statistics (DISES), Via Ponte Don Melillo, 84084, Fisciano (SA), Italy, riorio@unisa.it

[†] University of Salerno, Department of Economics and Statistics (DISES), Via Ponte Don Melillo, 84084, Fisciano (SA), Italy, rmdamore@unisa.it

[‡] University of Ferrara, Department of Economics, Institutions, Territory, Via Voltapaletto, 44100, Ferrara, Italy, sandrine.labory@unife.it

[§] University of Naples Federico II, Napoli, Italy, agnieszka.stawinoga@unina.it

1. Introduction

The importance of networks built by firms in order to create internal and external knowledge flows so as to create knowledge has been widely stressed, especially in high tech sectors.

From this perspective, the dynamics of firm growth and competitiveness results from the quality of interactions and coordination within their networks as well as from the interactions with external networks.

The way in which firms build and manage networks of research collaboration is an important issue to study in order to both gain a better understanding of innovation processes and to highlight policy implications, especially in a country like Italy, which is lagging behind in terms of innovation and high-tech sector development (European innovation scoreboards).

Many studies underlined that distance among agents is a crucial dimension in shaping research network: we try to increase the knowledge of research networks analysing how institutions deal with geographic and institutional distance: we observe, in the Italian "red" biotech sector, what kind of relationship exists between this two kinds of distance and we try to understand if such relationship is related to the nature of exchanged knowledge and to the stability of the interactions among agents.

The importance of the geographical extent of relationships in research networks has been widely stressed. According to Katz (1994) geographical proximity renders collaboration more likely, due to the scope to reduce the additional travel cost and time involved that are impediments to collaboration. Indeed this argument has become weaker, because of the improved and cheaper transportation possibilities and the rise of ICT, but this will not imply the so-called "death of distance", as geographical proximity still remains important for innovations for several aspects, relates to the face-to-face contacts (Evans, 2011). If geographic proximity is an advantage for the institutions involved in a collaboration process, geographic distance is conversely a barrier, therefore it implies a cost.

On the other side, high and specialized competencies may be found in distant partners, even outside national borders: overcoming geographic distance, in such a case, implies a benefit:

In the debate about its costs and benefits, the geographic distance has usually been considered with no reference to other kinds of distance. But recent studies have underlined that geographic proximity can be very important in a more indirect way, as it lets to overcome possible difficulties deriving from institutional distance. This term refers to the fact that relationships between different institutions will have to overcome possible differences in culture, norms, values and rules. Such kind of institutional distance typically arises in academic-company or academic-governmental collaboration, where different motivations and aims in doing research exist (Ponds et al 2007). According to Boschma (2005) geographical proximity can compensate for the lack of institutional proximity, and,

reversely, institutional proximity facilitates interaction over long geographical distances.

Such kind of relationship may be understood if we consider that, like geographic distance, institutional distance implies costs and benefits too: the differences in norms and values are barriers to overcome, therefore they imply costs; on the other side, the institutional variety allows firms to learn more knowledge, confronting varied knowledge bases. We may conclude that collaborating with "distant" partners may be fruitful, but costly; collaborating with a close partner may be less fruitful but cheaper: an economic choice has to be done. An easily understandable strategy for an agent is trying to obtain more benefits increasing one kind of distance in collaboration, reducing another kind of distance. Therefore a trade-off between geographic and institutional distance may exist.

In discussing the choice of innovative agents in terms of distance, the nature of exchanged knowledge cannot be neglected: many scholars, like Gittleman (2007), and Broström (2010), report that basic research is mainly conducted at an international level, while applied research is more locally focused. As concerns institutional distance, in the sector we are analysing, red biotech, institutional variety in applied research is necessary because of the necessary relations with users, which are in our case hospitals. However, relationships with users also exist at earlier stages of the R&D process, in order to orientate research towards specific users' needs (specific diseases, specific treatments). In fact a new feature of our research is to consider not only usual research institutions like universities and research centres, as Ponds et al. (2007) do, but also users, specifically hospitals, which are also involved in the research process of biotech firms.

In our analysis of geographic and institutional distance the stability of relationships is another aspect to keep explicitly into consideration. The above considerations about costs and benefits related to distance imply that collaborations may have different degrees of stability at different geographic scales: as a collaboration with a distant partner imply high costs, it will be rarely repeated (this is what D'Este and Iammarino (2010) observe about geographic distance). As a consequence of this, if we shift our analysis from the whole networks of collaborations to the more stable ones, the relationship itself between geographic and institutional distance may change.

We observe the phenomenon of research collaboration analysing the co-authorship of scientific publications in the Italian "red" biotech sector (biotechnology applied to life sciences).

The methodology is based on social network analysis and the use of specific indexes which, to our knowledge, have not been used in this type of research beforehand. The indexes are both the E-I index, that measures the homogeneity of links in a network, which we apply to our case adapting it to weighted networks. and the equivalence coefficient, that measures the frequency of links between two subjects, relative to their total number of links.

In addition, the results of the social network analysis are checked using regression analysis. This paper contains therefore innovations both in terms of content and methodology: the issue of the relationship between geographic and institutional distance is analysed taking extensively into account factors like the nature of the exchanged knowledge and the frequency of the collaborations; the

issue is studied using indexes already existing but used in different context and adapted to this purpose; a regression analysis is performed in order to complement the social network analysis.

The paper is structured as follows. The second section provides literature review and an outline of relevant questions to be explored in the paper's empirical research. The third section presents the data, while the fourth section explains the methodology. The fifth section presents the results of the empirical analysis, together with a synthesis of main results.. Lastly, the sixth section concludes.

2. Literature Review and Relevant Issues of Analysis

In the science-based sector relationships with universities or other public and private research centres are particularly important, in order for them to access the new knowledge that is constantly emerging (Gittleman and Kogut, 2003; Powell et al., 1996). Especially in the biotech sector the complexity of the knowledge base, together with the high uncertainty of research and the small size of most firms, induce them to create dense collaborative networks (Powell et al., 1996).

Giving the strong polarity of innovation activities (Lung, 1997; Puga 1999, Lallement et al 2001) many authors have stressed the importance of geographical proximity to collaborate, considering local proximity as the most favourable place for innovation through collaborations. The reduction of travel costs, the development in communication technologies have perhaps reduced the costs of collaborating with distant partners, but for many aspects geographic proximity remain important in the innovation process: tacit knowledge may be exchanged only with to face to face interaction and personal contacts at local level are easier (Katz, 1994; Liang and Zhu, 2002; McKelvey et al., 2003); geographic proximity results in higher spill-overs (Amin and Wilkinson, 1999; Jaffe et al., 1993; Anselin et al. 2000; Audretsch and Stephan, 1996).

On the other side, Malmberg and Maskell (2002) show that long-distance relationships allow higher variety in the knowledge base; Iorio et al. (2012) show that papers' citations, therefore scientific quality, rise with international collaboration. Therefore geographic distance may yield costs but they may be compensated by high benefits in terms of innovation, which might occur when long-distance relationships allow to both widen the knowledge base and collaborate with highly competent researchers.

This appealing and evolving framework explains the abundance of recent literature on the geographic distance. But in recent years, beyond the geographical dimension of distance, a great attention has been devoted to the institutional distance as a determinant key of creation of and diffusion of new knowledge. Boschma (2005) relates institutional distance to the fact that interaction are influenced, shaped and constrained by the institutional environment. Institutional proximity means that the transmission of knowledge may be facilitated by the presence of common institutional framework, therefore common culture, values and norms. Institutional distance therefore implies costs, that may be attributed to

control loss (the higher the institutional variety the more difficult it is to control the output of research) and to difficulties of coordination.

On the other side, the institutional variety allows firms to learn more knowledge (confronting varied knowledge bases) and to increase the speed of innovation (for instance, the direct contact with users like hospitals enable to orientate research to specific users' needs). Iorio et al. (2012) show that institutional variety appears to generate more scientific advance in the biotech sector (see Bonaccorsi and Thoma, 2012 too) [1]

We may conclude that both geographic and institutional distance imply costs and benefits: if a certain degree of substitutability between the two forms of distance exists, it is clear that decisions about the two kind of distance cannot be considered separately[2].

In fact Boschma (2005), who identified five kinds of proximity (cognitive, organisational, social, institutional and geographical) stated that the impact of one type of proximity must be examined in relation to the other types. Even Autant-Bernard et al (2007) evidenced that the proper impact of the geographical dimension must be more precisely assessed in relation to other types of proximity, such as institutional one.

Therefore many scholars analysed the existing relationship between institutional and geographic distance. According to Boschma (2005) geographical proximity can compensate for the lack of institutional proximity and *vice versa*. Ponds et al. (2007) show that research collaborations between different institutions is more likely over short distances, because the sharing of a common local labour market, the access to local funds and mutual trust induced by informal contacts and interaction helps overcoming the difficulties created by institutional distance. Thus researchers in firms and in universities have different incentive structures but geographical proximity helps to reduce their institutional distance.

Evans et al. (2011) show that academic scientists in the social sciences tend to prefer intra-institutional collaborations; when they seek for collaborators from other institutions, they tend to select collaborators in close geographic proximity: there is evidence of a trade-off between institutional and geographic distance for academic researchers.

However, an alternative view is possible. It may also be assumed that frontier research requires highly specialised competencies that are very dispersed, both from a geographic and an institutional points of view, therefore a firm doing research at the frontier might activate an international and heterogeneous network of collaboration. In contrast, if the knowledge is more 'ordinary', a firm might prefer activating a local and homogeneous network, so that we might expect local networks to be more homogeneous than international ones. If this view is correct, a

[1] See Bouba-Olga et al. (2012) for an exhaustive review of costs and benefits deriving from spatial and institutional distance

[2] Indeed different institutions are not perfect substitutes, as each of them has specific competencies, embodied in the single individuals and in the research groups. Therefore a partner is identified not only by its spatial and institutional distance but also (and perhaps mainly) by its peculiar knowledge. These considerations warns the scholars, who desire to study how the institutions that collaborate in scientific research deal with the spatial and institutional distance of the partners, from each easy generalization of the conclusion deriving from their observation; anyway it does not prevent them from searching some regularities.

trade-off between institutional and geographic distance should arise only among low-quality research projects, while among high quality projects a relation of complimentarity prevails. Iorio et al (2011) find some evidence of this kind of relationship, assuming papers' citations as a measure of project quality.

Other factors that may influence the relationship between institutional and geographic distance are the type of exchanged knowledge and the phase of R&D process.

Many scholars studied the relationship between phase of R&D and geographic distance. In the life science industry, geographical proximity has been shown to have different importance at different stages of the R&D cycle. In biotech, long-distance relationships are more likely to be focused on earlier stages of the R&D life-cycle (Feldman, 2000; Mariani, 2004; Gittleman, 2007), while in engineering geographically out-stretched linkages are more likely to be focused on the middle phases of the R&D life-cycle (Broström, 2010).

Other studies focused on the relationship between phase of R&D and institutional distance. Evans (2011) argues that academic researchers with industry ties are more likely to publish less, to divulgate their research and methods to a lesser extent. Firms are therefore more likely to collaborate in research with academia at less applied stages of research, more frontier knowledge, where commercial value is less immediate.

As mentioned above, institutional distance may imply a control loss; indeed, control loss is a particularly important problem for firms investing in R&D, since they risk not taking the returns from their invention. Institutional variety is therefore likely to regard early stages of the innovation process, namely basic research and frontier knowledge.

Gittelman (2007) tries to relate these aspects in a comprehensive view. She shows that biotech firms engage both in dense regional relationships and in long-distance relationships in their learning processes, and that local relations are more conducive to innovation than long-distance ones. The main reason for the higher effectiveness in terms of innovation of regional relationships is that they are more institutionally varied. Long-distance relationships tend to be less institutionally varied, because of both the costs involved in institutional variety and the nature of knowledge flows, since long-distance relations appear to be focused on basic research while short-distance ones concern more applied research. As a consequence, long-distance relationships are focused on basic stages of research where scientific impact is a key driver.

These considerations makes it clear that, in analysing the institutional-geographic trade-off, the nature of the knowledge has to be taken into account But going into details in the analysis of knowledge is extremely difficult, due to the particular characteristics of knowledge. Knowledge is heterogeneous, difficult to observe (especially tacit knowledge which is incorporated into thoughts and actions) and no model exist that can relate inputs and outputs in the knowledge creation process (David and Foray, 2001).

This paper makes a distinction, partly following Stoke (1994), between pure basic research, which does not have any practical purpose, and basic research

aimed at subsequent application fields. Biotechnologies are aimed at subsequent applications to improve human health, especially when developed in the biomedical field.

In biotechnology, the frontier between basic and applied research is blurred and the distinction between basic and applied research has to be made with caution. For instance, finding a new DNA sequence is basic research that however can have direct applications to the development of new medicines or new treatments. Precisely because the frontier is blurring, biotech firms are more likely to have interest in participating in basic research, possibly with other research centres or universities. We can expect that biotech firms will carry out more basic research in international networks, which they publish in scientific journals to raise their reputation in the scientific community, while applications of these basic findings will be studied more locally. In addition, in the biotech sector both basic and applied research might require institutional variety. At basic research level, collaboration with universities and research centres has been shown to be important for firms (Gittleman, 2005, 2007) in order for them to know the precise requirements in terms of new diagnosis or treatments. At applied research level institutional variety might also be important because firms test their new biomedical products collaborating with users, especially hospitals.

A last factor that has to be taken into account as potentially influencing the relationship between institutional and geographic distance is the frequency of interactions. D'Este and Iammarino (2010) find that the frequency of collaborations between university and industry is negatively related with the geographic distance of the partners. In other words, collaborations with institutionally distant partners happen more rarely if geographic distance is high too: as distance implies costs, there will not be a multiplication of costs with repeated interaction with distant partners. Therefore more frequent interactions will happen with closer partners, but this tendency may assume different intensity at different geographic scales; this may imply that the scale, even the direction of the institutional/geographic relationship may change if we consider different degrees of frequency of collaborations[3].

At the end of this literature review, we may summarise the main findings:

1. many scholars identify a trade-off between geographical and institutional distance;
2. basic research generally implies long-distance collaborations, applied research implies more local collaborations;
3. in many scientific fields, and particularly in the biotech sector, both basic and empirical research may require institutional variety;
4. the nature of the research (basic or applied) may influence the geographic/ institutional trade-off;

[3] Other factors may be identified as important in shaping the the geographic distance aspect of the collaboration network. For instance, D'Este and Iammarino (2010) themselves and Evans et al. (2011) underline the importance of research quality in shaping the geographic distance aspect of the collaboration networks.

5. more frequent collaborations involve little distance between institutions;
6. the frequency of collaborations may influence the geographic/ institutional trade-off.

In some cases these conclusions are not unanimously accepted (1) or have been little analysed (2, 5); in other cases the relationships, if existing, are not clear in their direction and intensity (3, 4, 6): the following empirical analysis has the aim to shed more light on them.

3. Data

We observe the phenomenon of research collaboration through the lens of co-authorship of scientific publications in the Italian "red" biotech sector (biotechnology applied to life sciences). Co-authorship networks have been widely used in the analysis of scientific collaboration in different disciplines (Barabasi et al., 2002; Newman, 2004; Powell et al. 1996). The biotech sector is particularly suitable for such kind of study, as it is characterized by a complex knowledge base, where the sources of expertise are widely dispersed and network relations are frequently used to access and to exchange this knowledge (Powell et al., 1996). In this science-based sector relationships with universities or other public and private research centres are particularly important, in order for them to access to the new knowledge that is constantly emerging (Gittleman and Kogut, 2003; Powell et al., 1996). In addition, most firms in the biotech sector – at least in Italy – are in the biomedical field, where relationships with users of innovation, namely hospitals, are particularly important, not only in the application phase but also in the research phase. In fact, many firms in our sample do collaborate with hospitals in the research phase and publish scientific papers together with researchers working in hospitals, in order to target research to the specific problems faced by this type of institutions.
Our focus on the biotech sector also allows us to assume that other dimensions of distance are small in our case. Researchers at both firms and universities, or research centres or even users such as hospitals can be assumed to share the same knowledge base, although with some slight differences, so that cognitive distance is very small. In addition, we assume that organisational and social distance are very small in this sector, since researchers are trained in the same 'epistemic communities' (Gittleman, 2007).
Given that cognitive, social and organisational distance can be assumed small, we can focus on geographical versus institutional distance.

The focus on firms' scientific publications inevitably excludes other forms of research collaborations, such as joint research not leading to scientific publications (but possibly a patent), joint research project financed by governments, exchange of researchers or other informal forms of collaborations (for instance, see Landry et al., 2007, for a review of the various forms of knowledge transfer between universities and industry). We are aware of this limitation but our focus allows us

to only consider collaborations yielding a scientific output, that is recognised in the scientific community thanks to the publication in scientific journal. However a large ratio of even formal collaborations ends in a joint paper (Gittleman, 2007)

Given the focus of the paper, the empirical analysis is performed at institutional level. We are interested in the relationships between institutions but we assume, like other works in the literature (for instance, Gittleman, 2007; Broström, 2010), that linkages are mediated by personal contacts. In other words, firm members relate to members of other organisations. Hence we use data on co-authorship of scientific papers and the institutions authors belong to.
More precisely, we record both the number of authors and the number of institutions for each paper and distinguish between the type of institution – using four categories, namely firms, universities, research centres and hospitals – and the geographical location of the institutions – with four levels, namely regional, Italian but extra-regional, European and extra-European. Our distinction between institutional types mainly results from the type of research conducted in the different institutions and from their peculiar goals: firms and hospitals conduct mainly applied research but firms have profits and hospitals have people's health as their specific goal; universities and research centres conduct both mainly basic research and both have the advance of knowledge as their goal, but diffusion of knowledge, through teaching at various levels, is specific of universities only. All academic institutions are included in the 'university' category, while all research centres which are not academic in their scope are included in the other category. Since private universities or research centres are very few in Italy, we do not distinguish between private and public academic institutions. The same applies to the hospital category where we include both public and private clinical institutions.

The data we use are all publications recorded in ISI-Web of Sciences where an Italian biotech firm is indicated in the authors' addresses. Our list of biotech firms is derived from a census realised by D'Amore and Vittoria (2008). We have 306 life-science for profit firms.

We obtained information about publications of the selected firms across the period 2003-2005. The record of each publication in ISI-Web of knowledge reports, among other kinds of information, the name of the authors and the name of the institutions the authors belong to. We extracted all the publications where the name of at least one of the selected firms (Italian life-science for-profit biotech firms) appeared among the institutions of affiliation.

115 of the considered firms made at least one publication recorder in ISI-Web of knowledge during the period 2003-2005. The total number of publications is 1053. The total number of the affiliation institutions of the authors is 900; there are 218 universities, 289 hospitals, 134 researcher centres and 114 other firms. The institutional co-operation in publication is very frequent: in 918 on the total number of 1053 publication (87.18%) the authors belong to more than one

institution (in the others 135 publications the only institution of affiliation is one of the biotech firms). The average number of institutions per paper is 3.43.[4]

Basing on this data, we built the network of co-autorship: the institutions the authors of the publications belong to are the nodes of the network.

In order to analyse the impact of spatial distance, we divided the papers in four categories: Italian regional papers (all the institutions the authors belong to are located in the same Italian region), Italian extra-regional papers (all the institutions the authors belong to are located in Italy, but at least two of them are located in different Italian regions), international European papers (at least one of the institutions the authors belong to is located in Europe, but there are no institutions located outside Europe) and international Extra-European papers (at least one of the institutions the authors belong to is located outside Europe). In our analysis we exclude papers written not in collaboration (written by authors belonging only to Italian biotech firms), therefore we have 918 papers. Among these, 149 (16.4%) are Italian regional papers; 401 (43,7%) are Italian extra-regional papers; 202 (22%) are international European papers; 166 (18.%) are international extra-European papers. Basing on this classification, we obtained four sub-networks, including all the papers of each categories.

The four described groups of papers are clearly scaled according to the geographic distance: in the regional group we have the smallest distance among the institutions; the distance increases in the extra-regional group, then in the European group; in the extra-European group there is the greatest distance. Of course, this is true on average, while some individual exceptions are possible: two institutions, to whom the authors of a paper belong, may be both inside one of the four territorial domains but may be more distant then two institutions located in two different domains.[5] Indeed, in our data the problem is negligible if we compare Italian regional and extra-regional groups of papers and if we compare European and extra-European groups, while it has some relevance only comparing the Italian extra-regional group and the European group; nevertheless, it is not so serious to limit the results of our analysis[6].

Besides, it has to be observed that in our distinction the national border has a role and this generate a linguistic, legislative, perhaps cultural characterization and homogeneity of each group of paper: all the papers of the regional and extra-regional group are characterized by the fact that Italian is the common language of all the authors (except the rare case of foreign authors belonging to Italian institutions), and that all the institutions are regulated by the same laws and perhaps by the same "culture"; in all the papers of the European and extra-European group the authors do not have Italian as common language (except the case, a bit less rare than the previous, of Italian scholars belonging to foreign

[4] A more detailed description of the biotech sector, of the data and more statistical information may be found in D'Amore, Iorio and Stawinoga (2010).

[5] For example, a paper written by authors belonging to the University of Milan and Palermo is an Italian extra-regional paper, while a paper written by authors belonging to the University of Milan and Zurich is an European paper: the geographic distance is higher in the first case, while, on average, the distances are smaller in the extra-regional domain then in the European one).

[6] We calculated, for each paper, the distance between the two most distant institutions whom the authors belong to. Considering a threshold of 600 kilometres, 58 extra-regional papers (14,46% of all extra-regional papers) are above this threshold, while 25 European papers (12,37% of all European papers) are under this threshold: considering a threshold of 700 kilometres, the figures are 16,83% (58 papers) and 10,22% (41 papers) respectively. If the distinction in territorial domains perfectly reflected a geographic scale, such figure should be 0 for at least on of the two groups. These figures demonstrate that such problem exists but it is not severe.

institutions) and Italian laws as regulative system, neither the culture inspiring individual behaviour is common.

These considerations induce us to consider as perfectly appropriate, to evaluate the effect of geographic distance, the comparison between regional and extra-regional papers and between the European and extra-European papers, while the comparison between extra-regional and European paper needs more caution.

Similarly the regression analysis does not adopt a continuous measure of the physical distance, in order to better compare the results of the regression with those of the social network analysis.

We also control for the research level (basic or applied) of the publication. The classification has been built by adapting a methodology developed by Lewison and Paraje (2004) (see Iorio, Labory and Paci, 2012, for details). We obtained 439 baisc papers and 339 applied ones; the other 140 papers are of a "mixed" nature.

4. Methodology

The starting point of the empirical research is the computation of an affiliation matrix $\boldsymbol{Z}$ of size (n × m), which is represented by a bipartite graph with two sets of nodes for institutions and publications and links connecting institutions and papers written by the institutions. In the matrix $\boldsymbol{Z}$ the generic element $z(i, j)$ ($i = 1, \ldots, n$; $j = 1,\ldots, m$) equals 1 if the paper j was written by the author affiliated to the institution i and 0 otherwise.

Subsequently, to obtain a collaboration network which is viewed as a social relationship between authors and can be represented by co-authorship in publication, an adjacency matrix $\boldsymbol{W}$ ($n \times n$) is calculated from the affiliation matrix $\boldsymbol{Z}$ by the product:

$$\boldsymbol{W}=\boldsymbol{Z}\boldsymbol{Z}^{\boldsymbol{T}} ,$$

The matrix $\boldsymbol{W}$ is an undirected weighted adjacency matrix, whose the value of the element $w(i,j)$ represents the number of co-authored papers for institutions i and j. If two institutions have no common publication the entries are equal to 0. The diagonal elements represent the total number of publications for each institution. If we are interested in taking into account only the presence and absence of ties we have to transform the matrix $\boldsymbol{W}$ in an undirected binary adjacency matrix $\boldsymbol{A}$ by setting all entries greater than zero to 1 and removing the diagonal elements.

The E-I index proposed by Krackhardt and Stern (1988) could then be computed. This index measures the relative homophily of a group while comparing the numbers of ties within groups and between groups and it is defined by :

$$\boldsymbol{E\text{-}I\ index = (E\text{-}I) / (E+I)}$$

where **E** (External) is the number of external ties (ties between nodes belonging to different groups); **I** (Internal) is the number of internal ties (ties between nodes belonging to the same group). The **E-I** index can be applied at three levels: the entire population, each group, and each individual. It ranges from -1 (all ties are internal, **E**=0) to +1 (all ties are external, **I**=0).

In our case the groups are the four kinds of institutions. A co-authorship relation between two hospitals indicates a collaboration within the same institutional group, therefore it is an internal link. A co-authorship relation between a firm and a university is a collaboration between two different institutional groups, therefore it is an external link. The value of E-I index is a measure of the mean propensity of each institutional actor to collaborate with a "different" actor rather than with a similar one. therefore it may be considered a suitable measure of institutional distance.[7]

The strength of collaboration among different institutions was analysed using the information about collaboration frequency and focusing on the weighted adjacency matrix ***W***. In addition to relational information a categorical attribute was associated with each node, based on the type of institution (universities, research centres, hospitals, firms).

The co-authorship frequency was normalised taking the number of publications of each institutions into account, in order to avoid basing the assessment of the strength (intensity) of collaboration between two institutions by only calculating the number of papers in common (co-authorship frequency). For this purpose a similarity index was used, namely the equivalence coefficient (EqC), defined by Michelet (1988). This index has been used to normalize frequency in co-word analysis (Polanco, San Juan 2006; Van Cutsem, 1994). According to the equivalence index, the degree of co-authorship between two institutions, i and j, is defined as:

$$EqC(i,j)=w(i,j)^2/w(i,i)*w(j,j).$$

The element *w(i,j)* represents the number of papers written by institutions i and j; the *w(i,i)* is the total number of publications of institution i. Since this coefficient is based on the product of conditional probabilities of the number of papers written by an institution knowing the number of publications of the other one, it has an easy interpretation in terms of probability theory.

The value of this coefficient can range from 0 to 1 and is maximized for pairs of institutions which collaborated in all papers they wrote. According to this index, a "weak" co-authorship relation between two institutions i and j exists if the value of *EqC(i,j)* is close to 0. A "strong" relation will be observed if *EqC(i,j)* is close to 1.

For the purpose of the joint analysis of the four important aspects of collaboration (spatial distance, institutional distance, "strength" of the links, content of the paper), we empirically investigate what happens to the different networks if the ties of different strength (from more occasional to more stable relations) are removed. The cut-off value is that of the equivalence coefficient,

[7] Institutional distance does not have a universally accepted measure: the literature generally assumes that there is institutional distance if two or more institutions of different kinds collaborate and there is no distance if the collaborations happens between similar institutions. In order to provide ranges of institutional distance beyond this 'yes or no' measure, this paper assumes that the institutional distance increases if, for each collaboration or for a network of collaborations, there is an increase in the ratio of heterogeneous linkages (between institutions of different types) over the total number of linkages. An index perfectly correspondent to our definition would be E/(E+I), while the E-I index is equal to (E-I)/(E+I). Nevertheless these two indexes are perfectly correlated, so for our purpose there is no practical difference in using one or the other. We preferred to use the E-I index because it is has been previously used in the literature.

starting with 0.1 and incrementing by 0.1. At each step, the networks related to matrix ***A*** have as many links as there are values in the matrix EqC, equal or greater than the cut-off value of the respective step. Then, basing on the binary network, we calculate the number of edges, the number of components the networks consist of and the E-I index. Finally, we compare the results of all cutting steps for all the networks.

Besides the complete networks of co-authorships, some sub-networks were computed, dividing the papers according to the localisation of the institutions and the content of the research (basic or applied)[8].

5. Results

A descriptive analysis of our data shows that the mean number of institutions involved in each paper increases with the geographic distance: it is 2.66 in the regional group of papers, 3.82 in the extra-regional, 3.68 in the European, 4.83 in the extra-European group; on the other side, the mean number, for each sub-networks, of how many kinds of institutions are involved in each paper is 2.13 in the regional group of paper; it is 2.42 for extra-regional papers; 2.24 for European papers; 2.49 for extra-European papers.

We are interested in the institutional distance and we measure it with the E-I index.. The value of the E-I index for the complete network of co-autorships is 0.346, indicating a predominance of "external" linkages (we recall that the index assumes positive values if heterogeneous linkages exceeds homogeneous linkages, up to 1 if all linkages are homogeneous; negative values, in the opposite case, up to -1, when all linkages are heterogeneous).

In order to analyse the relationship between institutional and spatial distance, our strategy is the following: we divided the publications in four groups, characterized by different levels of spatial distance, generating four sub-networks; then we calculated the global value of E-I for each of the four sub-networks. As it is possible to calculate the index for groups of nodes, we also obtained the E-I index for firms, that is the only fully represented group in our sample (we have all the papers including at least one Italian biotech firm in 2003-2005)

The following table shows the values of the weighted E-I index in the four sub-networks.

[8] The graph of the networks may be found in the appendix; for the main statistics of the networks, see D'Amore, Iorio and Stawinoga, (2011).

Table 1. Weighted E-I index for the three valued networks

	E-I index	E-I index for firms
Regional network	0.746	0.870
Extra-regional network	0,380	0.767
European network	0,325	0.493
Extra-European network	0,369	0.427

The value of the global index dramatically decreases as we move from regional to extra-regional sub-networks, indicating the existence of a trade-off between spatial and institutional distance: as geographic distance increases, the heterogeneity of the institutions involved in the co-authorships of the papers, hence the institutional distance, increases. However, this seems to hold clearly only comparing the lowest level of the geographic scale with the other three levels, which do not present large differences in institutional distance. The behaviour of the firms, expressed by the E-I index for firms, is more clearly addressed to the institutional-geographic trade-off, as the index always decreases with the enlarging of geographic scale.

Regarding the issue as to whether the nature of research may explain the trade-off between institutional and geographic distance, we find that the average institutional distance is higher in basic research than in applied (the weighted E-I index for the sub-network of basic research paper is 0.436, while it is 0.336 for the sub-network of applied research), but, notwithstanding an high concentration of basic research at a very local (regional) level, the relationship between geographic distance and content of research is not clear (we observe that 48.94% of the papers have a basic research content; the ratio of basic research paper is 59.57% at regional level, 43.18% at an extra-regional level, 52.76% at an European level and 49.07% at an extra-European level). Therefore there is not a clear relation between one kind of research and a combination of high institutional distance/low geographic distance (and between the other kind of distance and a combination of low institutional distance/high geographic distance).

A multivariate analysis is needed to explore this issue in depth. This is carried out with papers as statistical units, a measure of institutional distance as independent variable, the geographical level (dummy variables corresponding to the four geographic levels: region, extra-regional, European, extra-European) as dependent variables, the nature of research, also expressed with dummy variables (basic, mixed, applied), as control variables. Dummy variables for years are added: the world of scientific research and publication is rapidly and continuously changing, therefore even a quite limited period may imply systematic changes that could be controlled through year dummy variables.

As a measure of institutional distance, coherently with the previous analysis, we have the E-I index calculated for each paper. As this E-I index may assume continuous values from -1 to +1, it is possible to estimate an OLS linear regression

model. Because of the presence of heteroskedasticity, we estimate a model with robust standard errors.

The following table shows the results (coefficients with standard errors in brackets)

Table 2. Regression analysis (OLS with robust standard errors) on institutional distance

	Model 1	Model 2	Model 3 (only basic research)	Model 4 (only applied research)	Model 5 (E-I for firms only)
Geographic level					
Regional		0.3174*** (0.0632)			
Extra-regional	-0.1483*** (0.0519)	0.1688*** (0.0524)	-0.1213* (0.0662)	-0.1864* (0.0983)	-0.1905*** (0.0617)
European	-0.3170*** (0.0632)		-0.2070* (0.0796)	-0.5429*** (0.1217)	-0.2576*** (0.0716)
Etra-European	-0.3370*** (0.0615)	-0.0199 (0.0624)	-0.3099*** (0.0788)	-0.2369** (0.1057)	-0.3475*** (0.0709)
Research level					
Basic		0.1577*** (0.0405)			
Mixed	-0.2169*** (0.0614)	-0.0591 (0.0645)			-0.2176*** (0.0696)
Applied	-0.1577*** (0.0405)				-0.0744 (0.0529)
Year dummies	INCLUDED	INCLUDED	INCLUDED	INCLUDED	INCLUED
Firm dummies	EXCLUDED	EXCLUDED	EXCLUDED	EXCLUDED	INCLUDED
N.obs.	897	897	439	339	897
F statistics (degr.of freed.)	10.09*** (7,899)	10.09*** (7,899)	3,90*** (5,433)	6.50*** (5,333)	2.89*** (116,780)
R-squared	0.0743	0.0743	0.0400	0.0920	0.1963

Dependent variable: Models 1-4: E-I index for each paper; Model 5: E-I for firms for each paper
***significant at 99% **significant at 95% *significant at 90%

In Model 1 the excluded geographic dummy variable is the regional one and the results show that there is greater institutional distance at the regional than at the other three geographic levels, confirming the existence of the institutional/geographic trade-off. This result is statistically significant at a 99% level and holds *ceteris paribus*, that is even controlling for the nature of the research. Regarding the latter, applied and mixed (basic/applied) research imply less institutional distance than basic research (the excluded dummy variable) and this conclusion is statistically significant at 99%.

The estimation of the same model, but excluding other geographic dummy variables, lets to analyse if the geographical-institutional trade-off exists for all the geographic levels. The following results are significant at 99% level: regional papers have more institutional distance than extra-regional papers and extra-regional papers have more institutional distance than European papers; the

difference in institutional distance between European and extra-European level is not significant at 90% level. Therefore the trade-off clearly exists up to the European level. Comparisons between institutional distance at different research levels may be done with the same technique of rotating the excluded dummy variable: we may observe that there is no significant institutional distance between mixed and applied research. In the second column of Table 2 we present the results of the model obtained excluding the dummy variables for European level and for applied research (Model 2).

The same model was estimated separately for each research level too (clearly excluding research level variables, that are constant), resulting in the existence of the trade-off for each research level, although with slight differences in the significance (Models 3 and 4).

Analogously with the previous analysis with social network analysis, we also focused on firm behaviour, calculating, for each paper, the E-I index for the firms. We estimated a regression with this dependent variable, having the same dependent and control variables as Models 1, plus firm dummy variables: the behaviour about publications of different firms may be systematically differen, and firm dummy variables may be useful in controlling for these differences. The results of this model (Model 5) are also presented in Table 2, showing that the behaviour of firms toward institutional and geographic distance do not differ from the behaviour of all institutions wholly considered.

Further insights on the trade-off between institutional and geographic distance may be provided by considering the frequency of collaborations through the equivalence coefficient –The latter index is built as follows. If a paper has been written by authors belonging to two institutions (we call them A and B), an edge between A and B is built. If A and B have written more than a paper together (e.g 8 papers), the edge is weighted (the weight is 8). The equivalence coefficient calculates the relative weight of the edge with respect to the total number of edges (papers) of the two institutions. In other words, it calculates the "relative intensity" of the co-authorships between A and B. In what follows a low level of the equivalence index is interpreted as 'occasional' collaborations, while a high level denotes 'stable' collaborations.

Our strategy consists in analysing what happens to the value of the E-I index, in the whole network and in the four sub-networks, if we restrict the observation to the ties of greater strength (to stable collaborations).

We recall that the value of the E-I index for the complete network of co-authorship is 0.346. If we exclude the more occasional collaborations (edges with EqC equal or smaller than 0.1) the value of the E-I index decreases to 0.116, which means that stable relations are more homogeneous. In addition, the value of the E-I index constantly decreases and becomes negative as thresholds of the equivalent coefficient rise, indicating a predominance of homogeneous linkages among the most frequent collaborations (the index is -0.003 if we consider the edges above the 0.5 threshold; -0.089 above the 0.9 threshold)

We can therefore conclude that more stable collaborations happen among more homogeneous institutions, and *vice versa*.

This is consistent with the idea that institutional distance imply a cost: as collaborations among heterogeneous agents imply higher costs, they happen rarely, less frequently that collaborations among homogeneous agents.

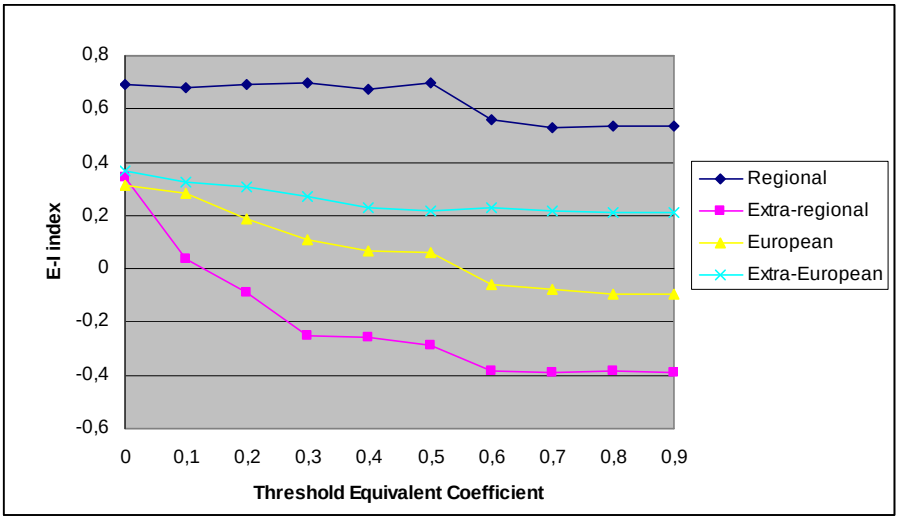
The same result is observed in each of the four sub-networks, but with a different intensity. At the extra-regional level the decrease is sharper (the E-I index decreases from 0.346 for all linkages to 0.038 for the edges above a value of 0.1 of the equivalence coefficient and becomes negative thereafter), while at the other levels the E-I decreases slowly.

Hence the interesting consequence that, comparing more stable collaborations, the strength trade-off between institutional and geographic distance becomes greater if regional and extra-regional are compared, but it disappears if other geographic levels are considered.

If the idea of an inverse relationship between the two kind of distance is grounded on the hypotheses that both distances imply costs and benefits, this results has perhaps the meaning that the logic of costs is less valid when repeated collaborations are considered: the repeated collaborations, without eliminating the barriers and the costs of overcoming them barriers, probably reduce both barriers and costs, reducing or eliminating the trade-off between different kind of distances.

Figure 1 shows the relationship between E-I index and different thresholds of equivalent coefficient, for the four geographically based sub-networks.

Figure 1: E-I index for different thresholds of equivalence coefficient



It is worth noting that 62.4% of collaborations are occasional (equivalence coefficient less then 0.1) in the complete network, which consists of 900 nodes and

1778 edges. Obviously, the number of edges remaining in the network reduces if we consider progressively higher thresholds of the equivalence index. The decrease is sharp from the complete network up to the 0.1 threshold, after which the decrease continues more slowly, becoming very limited after the threshold of 0.6.

This result is quite different in the four sub-networks: most collaborations are occasional (and homogenous) at the extra-regional level, while they are mainly stable at the other levels.

In the end of the analysis, we add that the reduction of E-I index when higher values of equivalence coefficient are considered happens both in applied and basic research papers, but with a stronger rate of decrease of the E-I index in the applied research.

5.1 Synthesis of the results

The main results of our analysis may be summarized in the following way:

-there is an inverse relationship between institutional and spatial distance (the institutional distance increases when the spatial distance decreases). This result is statistically robust comparing regional, extra-regional and European level; it is ambiguous comparing European and extra-European level;

-the basic research implies more institutional distance than applied research

-at a regional level there is a greater intensity of basic research than at the other geographic levels;

- the inverse relationship between institutional and geographic distance holds even if the nature of research is taken into consideration, that is “controlling” for it;

- the inverse relationship between institutional and spatial distance holds for both basic and applied research separately considered;

- there is an inverse relationship between institutional distance and the frequency of collaborations (the institutional distance decreases if we consider more frequent collaborations);

-the previous results holds at any geographical level; the magnitude of the effect is not identical at any geographical scale (it is much larger at the extra-regional level), implying some changes in the relationship between spatial and institutional distance if we consider more stable or more occasional collaborations.

6. Conclusions

In this paper we explore in depth what is the behaviour of the institutions involved in the scientific collaboration toward two dimensions of distance, spatial

and institutional, considering other two aspects: the nature of research and the frequency of the ties.

We analyse this issue thanks to a database of the co-authorship of scientific articles in the Italian "red" biotech sector in the period 2003-2005.

In shaping their research networks, the institutions make a balance of costs and benefits.

The several existing forms of distance among research partners imply a cost, but higher competencies (therefore higher benefits) may be found in distant partners. Therefore, supposing a certain degree of substitutability between potential research partners, a trade-off may exist between different kinds of distance. We provide a further enquiry on this topic, adding some novelties.

Firstly, we distinguish the institutions in four kinds (firms, universities, hospitals, research centres), while, in such kind of studies, usually only university-industry collaborations are analysed, or there is a distinction of the institutions in three categories (firms, universities, public research centres)

Another novelty of our research is that we keep the nature of the research, if basic or applied, into consideration. The existing literature induces us to think that both geographic and institutional aspect of collaborations are influenced by the nature of research. If this is true, an analysis of the relationship between geographical and institutional distance that does not consider the kind of research would be misleading. Thanks to a classification of the papers of our sample based on a criterion purposed by Levison and Parajee, we may verify if, in our sample, the nature of the research is correlated with the geographical and institutional distance; then we are able to include the nature of the research as a "control" in our analysis of the relationship between the two kinds of distances.

As a further step of our analysis, we suppose that there is a relationship between institutional distance and the frequency of collaborations: if, as we are supposing, the institutional distance implies a cost, collaborations will happen usually with similar partners, while collaborations among different partners will be more occasional. If this is true, the relationship between geographical and institutional distance may be different if we are considering more occasional or more usual collaborations.

Our study introduces some methodological novelties too. The analysis is conducted generating the network of co-autorships, then splitting this network in four sub-networks according to the geographical distance among institutions. As a measure of institutional distance we use an index previously used in different contexts and applications: the E-I index, a measure of homophily of the linkages among different groups; moreover, this index has been used in the binary networks, while we calculate it in a weighted network. As many dimension are included in the analysis, a multivariate analysis is needed, therefore the social network analysis is complemented by a regression analysis; the E-I, as a measure of institutional distance, is used in the regression analysis too.

In order to evaluate the relationship between the institutional distance and the relative frequency of collaborations, we combine the use of E-I index with an index, the equivalence coefficient, already previously used in the social network analysis but in different contexts.

We found a confirm of the existence of institutional/geographic trade-off, that does not depend on the nature of research, but it is valid both for basic and applied research. We also found that agents tend to collaborate more frequently with institutionally close partners, but this phenomenon has different strength at different geographic scales.

The results of the analysis are consistent with the view that both institutional and geographic distance are a cost, therefore institutions try to save money and/or efforts by substituting one with the other and avoiding too frequent, "distant" collaborations.

With this paper we desire to give a contribution to the understanding of how knowledge flows among innovative agents. We hope this contribution may be useful in designing proper policy measures in the field of scientific collaboration, that is so important for the competitiveness in the contemporary knowledge-based economy.

References

Amin A, Wilkinson F (1999): "Learning, proximity and industrial performance: an introduction", *Cambridge Journal of Economics*, 23, pp. 121-125

Anselin L, Varga A, Acs Z (2000): "Geographic and sectoral characteristics of academic knowledge externalities". *Papers in Regional Science*,79(4), pp.435–443

Audretsch, D., Stephan, E. (1996): "Company-scientist locational links: the case of biotechnology". *American Economic Review*, 86(3), pp. 641-652.

Autant-Bernard C., Billand P., Frachisse D., Massard N. (2007): "Social distance versus spatial distance in R&D cooperation: empirical evidence from European collaboration choices in micro and nanotechnologies", *Papers in Regional Science*, 86, pp. 495-520.

Barabasi, A.-L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., Vicsek, T. (2002). "Evolution of the social network of scientific collaborations", *Physics* A, 311(3-4), pp. 590–614.

Bonaccorsi A., Thoma G. (2007): "Institutional complementarity and inventive performance in nanotechnology", *Research Policy*, 36(6), pp. 251-283.

Boschma R. (2005): "Proximity and innovation: a critical assessment", *Regional Studies*, 39(1), pp. 61-74.

Bouba-Olga O., Ferru M., Pépin D. (2012): "Exploring spatial features of science-industry partnership: a study on French data", *Papers in Regional Science*, 91(2), pp. 355-376.

Broström A. (2010): "Working with distant researchers-distance and content in university-industry interaction", *Research Policy*, 39, pp.1311-1320.

D'Amore R., Iorio R., Stawinoga A. (2010): "The role of institutions and firm characteristics in the network of firm publications. An analysis of the Italian biotech sector through the social network analysis", *World Journal of Management*, 2(3), pp. 81-100.

D'Amore R., Iorio R., Stawinoga A. (2011): "Who and where are the co-authors? The relationship between institutional and geographic distance in scientific publications", *Working Paper INFER, 2011.4*

D'Amore R., Vittoria P. (2008): "Il nuovo settore delle imprese biotecnologiche. Fonti informative, indicatori statistici e ambiti di policy", *L'industria*, 2, pp. 329-346.

Dasgupta P., David P. (1994): "Toward a new economics of science", *Research Policy*, 13, pp. 3-20.

D'Este, P., Iammarino (2010): "The spatial profile of university-business research partnerships", *Papers in Regional Sciences*, 86, pp. 423-443.

Evans J.A. (2010): "Industry collaboration, scientific sharing and the dissemination of knowledge", *Social Studies of Science*, 40(5), pp. 757-791.

Evans T.S, Lambiotte R., Panzarasa P. (2011): "Community structure and patterns of scientific collaboration in business and management", *Scientometrics*,89(1), pp. 381-396.

Feldman M. (1994): *The Geography of Innovation*, Kluwer Academic Publishers, Dordrecht.

Gittelman M. (2005): "What makes research socially useful? Complementarities between in-house research and firm-university collaboration in biotechnology", *Revue d'Economie Industrielle*, 110(2), pp. 57-73.

Gittleman M. (2007): "Does Geography Matter for Science-Based Firms? Epistemic Communities and the Geography of Research and Patenting in Biotechnology", *Organization Science*, 18(4), pp.724-741.

Gittelman M., Kogut B. (2003): "Does good science lead to valuable knowledge? biotechnology firms and the evolutionary logic of citations patterns", *Management Science*, 49(4), pp. 366-382.

Iorio R., Labory S., Paci D. (2012): "The Research Collaboration and Publication's Citation Impact: Evidence in the Italian Biotech Sector", *Rivista Italiana degli Economisti*, 18(1), pp. 75-100.

Katz J. (1994): "Geographical proximity and scientific collaboration", *Scientometrics*, 31, pp. 31-43.

Krackhardt D., Stern N. (1988): "Informal networks and organizational crisis: an experimental simulation", *Social Psychology Quarterly*, 51(2), pp. 123-140.

Jaffe A., Trajtenberg M., Henderson R. (1993): "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *The Quarterly Journal of Economics* 108(3), pp.577–598.

Lallement R, Mouhoud EM, Paillard S (2007): “Polarisation et internationalisation des activités d’innovation: Incidences sur la spécialisation technologique des nations” in Rallet A, Torre A (eds) *Quelle proximité pour innover?*, L’Harmattan, Paris.

Landry R., Amara N., Ouimet M. (2007): “Determinants of knowledge transfer: evidence from Canadian university researchers in natural sciences and engineering”, *Journal of Technology Transfer*, 32, pp. 561-92.

Lewison, G., Paraje, G. (2004): “The classification of biomedical journals by research levels, *Scientometrics*, 60(2), pp. 145-57.

Liang L., Zhu L. (2002): “Major factors affecting China’s inter-regional research collaboration: regional scientific productivity and geographical proximity”, *Scientometrics*, 55(2), pp. 287-316.

Lung Y (1997): *Organisation spatiale et co-ordination des activités d’innovation des entreprises*. Rapport pour le Commissariat Général du Plan, Paris.

Malmberg A., Maskell P. (2002): “The elusive concept of localization economies: towards a knowledge-based theory of spatial clustering”, *Environment and Planning A*, 34(3), pp. 429-449.

Mariani, M. (2004): “Research Collaborations among Inventors and the Location of R&D in the European Chemical Industry”, in: F. Cesaroni, A. Gambardella & W. Garcia-Fontes (eds.), *R&D, Innovation and Competitiveness in the European Chemical Industry*, Kluwer Academic Publishers, Dordrecht.

McKelvey M., Alm H., Riccaboni M. (2003): “Does co-location matter for formal knowledge collaboration in the Swedish biotechnological-pharmaceutical sector?”, *Research Policy* 32, pp. 483-501.

Michelet B. (1988): *L’Analyse des Associations*. Thèse de doctorat, Université de Paris 7.

Newman, M.E.J. (2004): “Coauthorship networks and patters of scientific collaboration”, *Proceedings of the National Academy of Sciences*, 101, pp. 5200–5205.

Polanco X., San Juan E. (2006): “Text data network analysis using graph approach”, *Proceedings of the 1st International Conference on Multidisciplinary Information Sciences and Technologies* (InSciT2006), vol. II, pp. 586–592.

Ponds R., van Oort F., Frenken K. (2007): “The geographical and institutional proximity of scientific collaboration networks”, *Papers in Regional Science* 86(3), pp. 423-443.

Powell W., Koput K., Smith-Doerr L. (1996): “Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology”, *Administrative Science Quarterly*, 41, pp.116-145.

Puga D (1999) “The rise and fall of regional inequalities”, *European Economic Review* 43, pp.303–334

Van Cutsem B. (Ed.) (1994). *Classification and dissimilarity analysis* (LNS 93). Springer.

Appendix

The following figures (Figures A1, A2, A3, A4) show the graphs of the regional, extra-regional, European and extra-European networks of co-autorships. Different shapes and colours represent the different kinds of institutions: red triangles represent firms; yellow circles represent universities; green squares represent hospitals; blue squares represent research centres.

Figure A1 Regional network

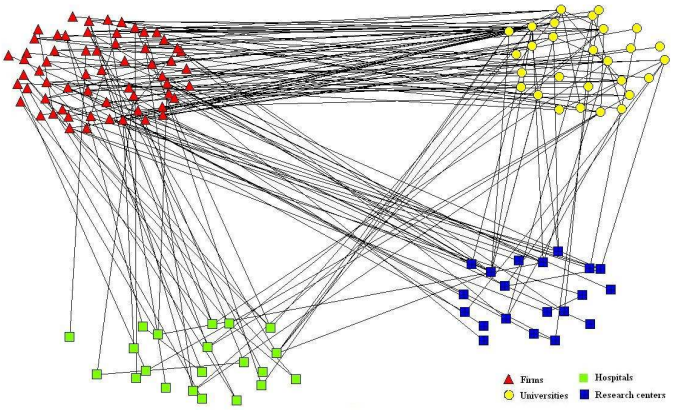


Figure A2 Extra-regional network

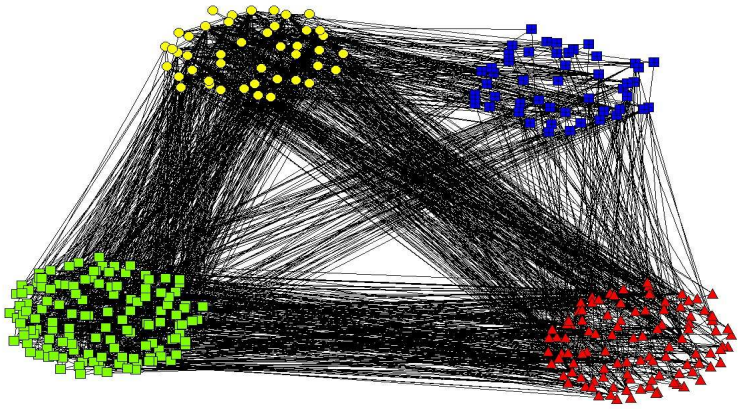


Figure A3 European network

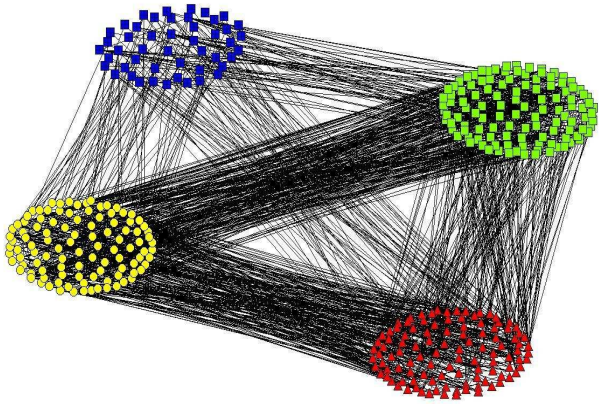


Figure A4 Extra-European network