DiSES Working Papers

Università degli Studi di Salerno
Dipartimento di Scienze Economiche e Statistiche
>>>**www.dises.unisa.it**

How Do the Institutions Involved in Scientific Collaboration Deal With Different Kinds Of Distance? An Analysis Of The Co-Autorships Of Scientific Publications

Rosamaria D'Amore
Roberto Iorio
Sandrine Labory
Agnieszka Stawinoga

Working Paper No. 3.222

ISBN: 978-88-6197-055-7
ISSN: 1971-3029

Dipartimento di Scienze Economiche e Statistiche
Università Degli Studi di Salerno
Via Ponte Don Melillo – 84084; Fisciano (SA) – Italy

Tel +39-089-96.21.55
Fax +39-089-96.20.49
E-mail dises@unisa.it
Web www.dises.unisa.it

HOW DO THE INSTITUTIONS INVOLVED IN SCIENTIFIC COLLABORATION DEAL WITH DIFFERENT KINDS OF DISTANCE? AN ANALYSIS OF THE CO-AUTORSHIPS OF SCIENTIFIC PUBLICATIONS

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

Abstract. In the scientific collaborations among different institutions at least two types of distance may be identified: spatial and institutional. We talk about spatial distance if the institutions are located in different places; about institutional distance if the collaborating institutions are of a different kind.

Given that there are both barriers and incentives to overcome such distances, the specific focus of this paper is on the way the institutions involved in the collaboration deal with them. We are interested in studying if more stable collaborations have different characteristics from the more occasional ones on the point of view of spatial and institutional distance,. We also aim to study also what kind of dependencies exist among these types of distance: if there is a trade-off among them, or a relation of complimentarity. Moreover, we take the content of the research, if basic or applied, 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) usually used in different applications.

This kind of analysis may highlight how knowledge flows among innovative agents and should be taken into consideration by the policy maker that aims to promote research collaboration between different institutions.

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

JEL Classifications: L65; O33; O31

Notes and Acknowledgements. This paper derives from a strict collaboration among the authors, both on the conceptual and empirical point of view. Anyway, sections 2 and 5 (except last paragraph) have been written by Rosamaria D'Amore; sections 3 and 7 by

[*] University 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

Roberto Iorio, section 4 and the last paragraph of section 5 by Sandrine Labory; section 6 by Agnieszka Stawinoga. Introduction and conclusions have been jointly written.

For many precious suggestions we are grateful to Maria Rosaria D'Esposito, Giuseppe Giordano and two anonymous reviewers.

1. Introduction

In the modern knowledge-based economy the innovation usually derives from collaboration of different agents, often located in different regions or countries and often belonging to different institutions (OECD, 1996). A rich literature underlines on one side the ease given by modern information technology to distant collaborations, on the other the persistent importance of vis-à-vis collaboration (Katz, 1994; Liang and Zhu, 2002; McKelvey, Alm and Riccaboni., 2003). Other scholars studied the difficulties deriving from collaboration between different institutions, particularly between university and industry, which have different research goals and incentive structure (Dasgupta and David, 1994); nevertheless collaboration among different innovative institutions may be fruitful in terms of quality, because of the existence of complementarities (Bonaccorsi and Thoma, 2007; Iorio, Labory and Paci, 2012).

In other words, there are two kinds of distance, spatial and institutional, that represent a form of, more or less high, barrier to the collaborations, not completely overcome by the chances given by the modern technologies and by the incentives given by the benefits deriving, as reported above, from the cooperation with distant subjects.

This framework becomes even more complex if we consider that different institutions are not perfect substitutes, but 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. This last consideration does not prevent however from searching some regularities in the behaviour of the institutions involved in scientific collaborations toward the managing of spatial and institutional distance, that is the goal of our study; moreover, supposing that the institutions do not consider the two kinds of distance separately, but take their interaction into consideration, we also try to shed some light on the theme of the relation between the two dimensions of the distance. Given this articulated theoretical framework, an all-inclusive view may be misleading. Our idea is that it is needed to distinguish the collaborations according to their degree of "stability": we suppose that the choices of the institutions, under the aspect of distance of the partners, may be different for occasional and usual collaboration. Another distinction we suppose is needed according to the nature of knowledge involved in the common research. This paper goes in fact in depth into the analysis of the nature of the knowledge flows between institutions, distinguishing between basic and applied research.

We observe the phenomenon of research collaboration through the lens of co-autorship of scientific publications in the Italian "red" biotech sector (biotechnology applied to life sciences) . The co-authorship networks has been widely used in analysis of scientific collaboration in different disciplines (Barabasi et al., 2002; Newman, 2004; Powell, Koput and Smith-Doerr, 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, Koput and Smith-Doerr, 1996).

Basing on a database including all the publications done by the Italian biotech firms from 2003 to 2005, we build the network of co-autorships, where the institutions the authors of the publications belong to are the nodes of the network. They are classified in four categories (firms, universities, hospitals and research centres) and their localization is registered too.

The main instrument of analysis are two indexes born in the context of the social network analysis, but previously adopted for different topics: the E-I index, that measures the homogeneity of links in a network (we consider a link homogeneous if a co-autorship happened between authors belonging to the same kind of institution, heterogeneous otherwise) and the equivalence coefficient, that measures the frequency of links between two subjects, relatively to the total number of links of the same subjects.

This paper contains therefore both an innovation in term of theoretical content and a methodological innovation, given by the application of such indexes to a context where they are not usually used.

The paper is structured in the following way: after this introduction, in the second section there is a review of the more relevant literature for our theoretical premises in the third section some considerations about costs and benefits deriving from distance are presented; in the fourth section we present some characteristics of collaborations and joint publications in the biotech sector; the fifth session presents the used database and a descriptive analysis of the data; in the sixth session the methodology used for the analysis is illustrated and the seventh section contains the empirical analysis (the results of the analysis and a brief discussion); some final considerations conclude the paper.

2. The relevant literature

The economic studies about innovative networks mainly focus on the networks involving firms and particularly analyse the behaviour of the firms. The focus of this paper is not concentrated on the firms, but this perspective is not inadequate, as we have to consider the peculiar nature of our data, that are papers involving at least one biotech firm.

The importance of networks between innovative agents built by firms in order to create knowledge flows within and outside the firms, so as to create knowledge, has been widely stressed, especially in high tech sectors.

A number of studies have shown the structure of firms' innovative or productive networks, particularly focusing on the territorial aspects of collaborations, often related with other relevant dimensions of the collaborations; Wagner and Leydersdorff (2005) report that there is evidence of the growth of international collaboration in science; the authors explain it using the principle of preferential attachment: they suggest that international collaboration arises as a self-organising phenomenon, whereby the selection of partners and the location of the research rely upon choices made by the researchers themselves rather than emerging through national or institutional incentives or constraints. This interpretation implies that the networks emerging in specific disciplines should be relatively stable through time.

Autant-Bernard et al. (2007) analyse the role of geographical distance and of the "network effects" (the position and role in the network of collaboration) in affecting the probability to collaborate in R&D projects. They also take the frequency of collaborations into consideration: they find that, among the firms that are involved in many projects, there is no evident influence of spatial distance on the probability to collaborate, while there is a clear influence of the firm's position within the network (number of direct and

indirect partners; social distance between firms); if firms involved in one single project are taken into account too, both geographical distance and social network effects matter.

Scherngell and Barber (2011) find that the spatial proximity increases the probability to collaborate between different organizations, but other factors may act in the same way: the thematic distance, the experience in projects of the same kind, the prior acquaintance and the centrality of the institution in the network of collaborations.

Other papers explicitly analyse the theme of relationship between spatial and institutional distance. McKelvey et al. (2003), studying, as we do, biotechnology in a national context (Sweden), find a trade-off between spatial and institutional distance: geographical co-location is more important for inter-institutional collaboration (firms with universities) than for collaboration among the same kind of institutions (firms with firms; universities with universities). Ponds, Van Oort and Frenken (2008) argument that both spatial and institutional distance imply a cost; as agents try to minimise costs, there is a trade-off between the two kinds of distance: the more is the spatial distance, the less is the institutional distance and *vice versa*; therefore local networks should be more heterogeneous than international networks. They observe the co-autorships in scientific publications, in eight technological fields and their conclusion is consistent with their hypothesis.

D'Amore, Iorio and Stawinoga (2012), analysing the same database considered in this paper, sustain that another line of argument is possible: highly specialised competencies are very dispersed, on a geographic and institutional point of view, therefore a firm that needs on-the-frontier knowledge activates an international and heterogeneous network of collaboration; if, on the contrary, the required knowledge is more ordinary, a local and homogeneous network may be activated. If this argument is correct, local networks should be more homogeneous than international networks. Indeed those authors suppose that the effects indicated by the two hypotheses (trade-off and complimentarity) coexist and the predominance of one or the other may depend by other factors, like the quality and relevance of the scientific project or may be different for different kinds of institutions. In fact they find that the relationship between spatial and institutional distance is inverse among rarely cited papers (supposing they derive from research project of limited quality or relevance), while the relationship is direct among frequently cited papers (deriving from research projects of high quality or relevance). They also find that different institutions have different behaviour and attitudes in managing the balance between the two kinds of distance.

Other papers analyse the relation between spatial distance and the content of research: Broström (2010) finds that geographical proximity is important for short-term projects of a very applied nature, because the exchange of tacit knowledge is particularly relevant for this kind of research, while in long-term projects it is generally easier to work across geographical distance.

3. Costs and benefits of "distant" partnerships

Ponds, Van Oort and Frenken (2008) sustain that both institutional and spatial distance imply a cost: collaborating with a distant partner (in institutional or geographic terms) is more expensive than collaborating with a near one: the collaboration with a physically distant partners implies direct costs, given by transportation and communication costs (even though such costs are decreasing) and the loss of the benefits given by the face-to-face contacts, particularly important for the transmission of tacit knowledge; the costs deriving from the institutional distance are given by the differences in goals, rules and values. On the other side, collaborating with an international or a different (in institutional terms) partner may give more benefits than a local or similar one, because the formers

may own that knowledge not owned at a local level or among institutions of the same kind (D'Amore, Iorio and Stawinoga, 2011). There are therefore negative (on the front of the costs) and positive (on the front of the benefits) incentives to collaborate repeatedly with more heterogeneous and distant partners.

Indeed, looking more in depth on the front of the costs, the collaboration with a distant partner rather than with a close one may imply an higher "cost of entry", to establish a collaboration, but, once it is activated, the cost to repeat the cooperation (the "marginal cost") may be not significantly different.

We do not have enough theoretical notions or empirical studies to precisely establish such differences in terms of benefits and costs; anyway such considerations about costs and benefits may address the empirical analysis and help to interpret the results. In fact, it may be useful to assume the very simple framework of a profit maximising agent, where the collaborations are the inputs of a "production function"[1]. A very simple model will consider four kinds of collaboration (spatially close and distant; institutionally close and distant) and each innovative institution has to decide the optimal number of each kind of collaborations to activate, that is the optimal value of the inputs. Each collaboration increases the total benefits of the firms but implies a cost. The balance between benefits and costs (more precisely, in the "neo-classical" view, between marginal benefits and costs) determines the optimal solution. It is of course possible that, for different kinds of distance, the balance between benefits and costs is different (it is, for instance, possible that an increase in spatial distance increases the benefits rather than costs, while the opposite happens for institutional distance, generating different optimal solutions). Besides, it is possible an interaction between the two kinds of distance, therefore an interdependence among the optimal solutions for each of them; the literature review of the previous section (see Ponds, Van Oort and Frenken, 2008 and D'Amore, Iorio and Stawinoga, 2011) reported some argument regarding the twofold possible relations between these two kind of distance. Finally, the optimal solutions may depend by several "exogenous" factors (the parameters of the problem): we consider if and how the distinction between basic and applied research is one of these factors, therefore if it has an effect on such results.

As we deal with aggregate values and we have different scales of collaborations (some institutions have few collaborations, some others have many), we do not consider the absolute number of collaborations for each institutions but a relative measure respect to their scale of collaboration, that we will illustrate later.[2]

4. Research collaboration and joint publications in the biotech sector

We study the way innovative agents deal with spatial and social distance, focusing on the co-autorships of scientific publications in the biotech sector, and within it, the "red"

[1] Of course we keep in mind that the substitutability among the institution is far from perfect, because, as reported previously, each of them has its peculiar knowledge. Besides, the linkage between input and output of the research is not deterministic. Anyway this "neo-classical" view may correspond some really existing driving forces, particularly in an aggregate view, therefore it may help in understanding the phenomenon under analysis.

[2] Let us suppose that an institution A has 9 collaborations with close partners and 1 collaboration with distant partners; an institution B has 15 collaborations with proximate partners and 5 collaborations with distant partners and an institution C has 40 collaborations with proximate partners and 60 with distant partners. Considering the absolute values of collaborations, we should conclude for a prominence of collaborations with distant partners (66 against 64). But the percentage of collaborations with proximate partners is 90% for A, 75% for B and 40% for B; with distant partners is 10% for A, 25% for B and 60% for B: considering the "intensity" of collaborations, we may conclude that this is overall stronger for proximate partners rather than distant ones. Indeed, we will use a more sophisticated measure of the "intensity" of collaborations.

biotech, namely biotechnologies with biomedical applications. Moreover, we choose a specific national context, Italy.

The biotech sector is particularly suitable for a study about research collaborations involving different institutions, because it relies mostly on inter-organizational collaborations. Powell, Koput and Smith-Doerr (1996) show that networks of collaboration ventures serve as a primary institutional arrangement governing exchange and production in this sector. The main reason why networks are so pervasive in this sector is also shown to be its complex and expanding knowledge base, implying that firms' learning processes arise in networks rather than within the firm boundaries. The relevance of networks in biotech sector is confirmed by the abundance of studies using the social network analysis in this sector: besides the paper cited above, we remind at least Powell (1996) and Gay and Dousset (2005).

The new knowledge generated by the collaborations not only takes the form of industrial innovations, but it is often disclosed trough the scientific publications: research collaborations often generate co-authored publications. Over two-thirds of even formal alliance partners in this field also appear as partners in scientific publications (Gittelman, 2005) and there is a close link between successful patents and scientific publications (Gittelman and Kogut, 2003; Murray and Stern, 2007).

Therefore, if the aim is to study the characteristics of the knowledge exchanges inside a technological field, considering that data on publications are usually of high quality and easy to access, it is possible to study the publications of the firms.

The sector we choose is thus particularly interesting for distinguishing type of knowledge and type of networks, because the organisation of the innovation process in this case is not straightforward, in that it is not clear how activities aiming at understanding fundamental problems and activities orientated towards practical problems are distinguished and organised in specific networks. An exploratory study of the kind we suggest appears to be therefore useful.

Regarding the geographical context of analysis, although not a leader in this sector, the Italian biotech sector has been constantly growing in the last decade and Italian biotech firms have been able to build global innovation networks. Iorio, Labory and Paci (2012) show that Italian biotech firms' networks extend both geographically and institutionally, as relationships extend in the national territory and abroad, and as firms build relationships with different institutions, including universities, other research centres and hospitals.

5. The data

In order to build a database of scientific publications in the biotech sector we made an intersection of two databases: *i*) RP Biotech data base; *ii*) ISI Web of Science. They are briefly described in the following.

RP Biotech data base. It is a collection of potentially all the Italian firms belonging to the biotech sector, active at December 2005. In this study we considered only the 306 life-science for profit firms.

ISI databases, especially the Science Citation Index®, and the web-based version Web of Science, is a detailed bibliometric database of journal articles and citations of worldwide research literature, that contains 14.000 international peer-reviewed scientific and technical journals.

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. Then we identified five categories of institutions (universities, research centres, hospitals, Italian life-science for-profit biotech firms, other firms) and established which category each institution belongs to. In the analysis presented in this paper, all firms are considered together, therefore we have four institutional categories.

115 of the considered firms made at least one publication during the period 2003-2005. The total number of publications is 1053. The total number of the affiliation institutions of the authors is 900; besides the 115 Italian life-science biotech firms, we identified 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.[3]

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 two categories: national papers (all the institutions the authors belong to are Italian) and international papers (at least one of the institutions the authors belong to is non Italian). 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, 550 (60.57%) are national papers, 362 (39.43%) are international papers. Basing on this classification, we obtained two sub-networks, one including all the national papers, one including all the international papers.

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), who classify papers into applied and basic using a filter (list of words) of articles' titles; we prefer to classify the papers using their subjects, because, although it is a more general categorisation of papers than the specific words in title, they provide a fairly adequate idea of whether the research is applied or basic. We do classify all the publications of biotech firms, being them biomedical or not. The choice of filter is arbitrary to a certain degree: we review different classification method (Narin and Hamilton, 1996) and analyse the single terms contained in the subject in order to establish our filter (see Iorio, Labory and Paci, 2012, for details of the filter). Excluding the papers that it was not possible to classify, the papers that are in a mixed category, between basic and applied research and the papers written without collaboration, that we excluded from the analysis, we obtained 439 papers of basic research (259 are national papers, 180 are international papers) and 339 of applied research (213 are national papers, 126 are international papers). We were therefore able to build the two sub-networks of basic and applied papers, and, crossing the information about location and nature of research, the four sub-networks (national basic, national applied, international basic, international applied).

6. The methodology of the analysis

In order to explore more in depth spatial and institutional distance and to be able to consider other two aspects such as the kind of research and the "strength" of collaboration in the scientific collaboration of Italian biotech sector, we propose the following strategy.

[3] 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).

In order to represent the data as a co-authorship network we start with 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.

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

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

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 publication in common 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.

In order to analyse the strength of collaboration among different institutions we take into account the information about frequency of collaboration and we focus further analysis on the weighted adjacency matrix $\boldsymbol{W}$. In addition to relational information we associated a categorical attribute with each node: the type of institution (universities, research centres, hospitals, firms).

In our opinion, because it is difficult to assess the strength (intensity) of collaboration between two institutions by calculating simply the number of papers in common (co-authorship frequency) but it depends also on how many publication each institution wrote, it is necessary to normalize the co-authorship frequency. In order to do it we propose to use a similarity index, namely the equivalence coefficient (EqC), originally defined in 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:

$$\boldsymbol{EqC(i,j)=w(i,j)^2/w(i,i)\cdot w(j,j)}.$$

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. The value of this coefficient can range from 0 to 1 and it is maximized for pairs of institutions which collaborated in all papers they wrote. According to this index, we indicate a "weak" co-authorship relation between two institutions i and j 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 paper we investigate the issue of homophily by analysing the degree of homogeneity of co-authorship relations among different institutions. In this part of analysis the institutional distance will be seen as the heterogeneity of relations among different institutions in the network.

In order to measure this aspects of relational system, we focus on the E-I index proposed by Krackhardt and Stern (1988). 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-autorship relation between two hospitals indicates a collaboration within the same institutional group, therefore it is an internal link. A co-autorship relation between a firm and an 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.

As described above, besides the complete networks of co-autorships, we built some sub-networks dividing the papers according to the localization (national or international) of the institutions and the content of the research (basic or applied)[4].

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 defined by the value of equivalence coefficient, starting with 0.1 and considering the increment of 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.

7. The results of the empirical analysis

As illustrated above, the basic instrument of our analysis is the equivalence index. 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, respect to the total number of edges (papers) of the two institutions; in other words, it calculates the "relative intensity" of the co-autorships between A and B (in the following, when this index is low, we talk of "occasional" collaborations, when it is high we will talk of "stable" collaborations).[5]

Our strategy consists in analysing what happens to the different networks if we restrict the observation to the ties of even greater strength.

With regard to the complete network of 900 nodes and 1778 edges we observe that 62.4% of collaborations relations are weak and can be considered occasional because the edges has the value of the equivalent coefficient smaller than 0.1 (Figure 1). Obviously, the number of edges remaining in the network reduces if we consider progressively higher

---

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

[5] An example may be useful: if, given that A and B (authors belonging to A and B) have written 8 papers together, A (authors belonging to A) have written 10 papers overall, B (authors belonging to B) have written 20 papers overall, the edge between A and B has an equivalence coefficient of $8^2/10*20 = 0.32$.
In this example A have written the 80% of its paper with B; B have written 40% of its papers with A. If these percentages increase (the "relative intensity" of collaborations between A and B increase), the equivalence coefficient increases: if A and B have written 7 papers together, but A have written 8 papers overall (therefore 87.5% of its paper with B), B have written 10 papers overall (therefore 70% of its papers with B), the equivalence coefficient is $7^2/8*10 = 0.61$. The example clearly shows that the equivalence coefficient depends by the relative weight of the collaboration, not by the absolute weight (in fact, in our example, it increases even if the number of the collaborations decreases, from 8 to 7).

thresholds of the equivalence index. 18.9% of the edges are above the threshold of 0.3; 15.1% of the edges are above the threshold of 0.5; 8.6% of the edges are above the threshold of 0.7; 8.3% of the edges are above the threshold of 0.9. Therefore, we have a strong decrease when we pass from the complete network to the threshold of 0.1 (more then half of the collaborations are removed, being very occasional), then the decrease continues more slowly, becoming very limited after the threshold of 0.6.

But, as the aim of the analysis is to evaluate how the agents deal with spatial distance, we conduct the same kind of analysis in the two sub-network of national and international collaborations.

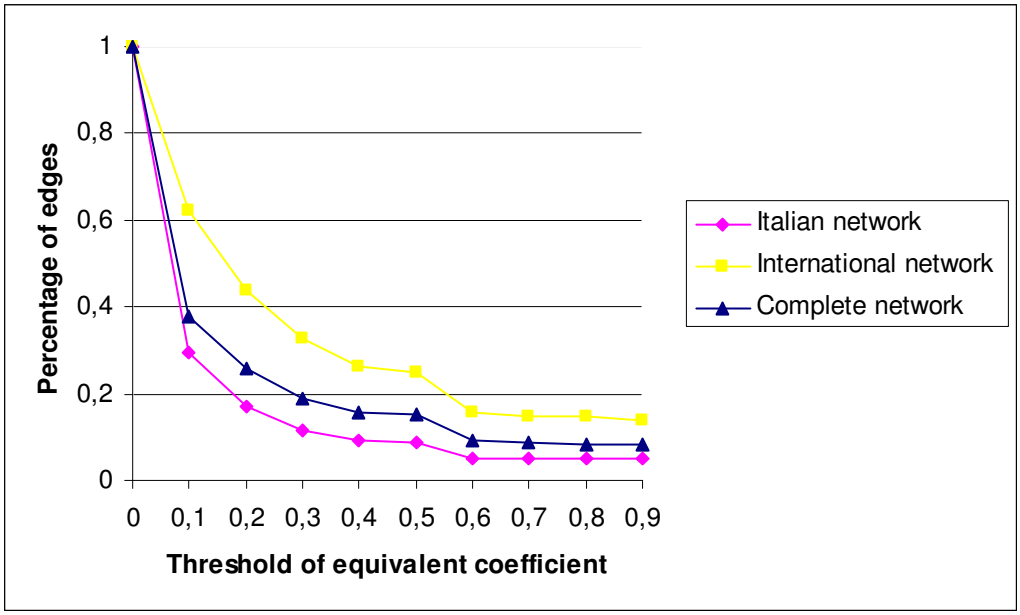
We observe that, in the national network (379 nodes, 1847 edges), only 29.6% of the collaborations are above the threshold of 0.1, while in the international network (668 nodes, 3079 edges) the percentage of collaborations above that value is 62.1% (Figure 1). Even considering higher thresholds, the percentage of edges that are above is always higher for the international network, even if the difference between the two networks reduces. If we consider the threshold of 0.9, in the national network 5% of the collaborations are so intense to be above that threshold; this percentage is 13.8% in the international network.

This comparison remarks an interesting difference, that let to obtain the first main conclusion:

1. a large majority of the collaborations involving only Italian institutions are occasional, while the collaborations involving at least one international institution are significantly more "stable".

Figure 1 shows, for the various thresholds of the equivalence coefficient (on the horizontal axis), the percentage of edges that are above that threshold (on the vertical axis), and this is shown for the complete network of co-autorship (blue line), the national sub-network (purple line) and the international sub-network (yellow line).

Figure 1- Percentage of edges above thresholds of equivalence coefficients: complete network, national and international sub-networks



Another aspect of our analysis concerns the "institutional distance", calculated through the E-I index. We recall that the index goes from -1, indicating a complete homogeneity of linkages (each institution has only "internal" linkages: it collaborates only with institutions of the same kind), to +1, indicating a complete heterogeneity of linkages (each institution has only "external" linkages: it collaborates only with institutions of a different kind).

The value of such index for the complete network of co-autorship is 0.346, indicating a predominance of "external" linkages.

This value is calculated considering all the co-autorships of our database. Let us now analyse what happens if we restrict our analysis to the network after removing weak relations. 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 stronger relations are more homogeneous. Indeed, the value of the E-I index constantly decreases if we consider even higher thresholds of the equivalent coefficient, becoming negative (it is -0.003 if we consider the edges above the threshold of 0.5; it is -0.089 if we consider the edges above the threshold of 0.9) (Figure 2).

This result lets to obtain the second main result:

2. more "stable" collaborations happen among more homogeneous (of the same kind) institutions; more occasional collaborations happen among more heterogeneous (of different kind) institutions.

Let us now move to the further point of our interest, that is the relation between the spatial and institutional distance; we obtain a measure of it comparing the value of the E-I index in the two sub-network, national and international. If we compare the values considering all the papers, no significant difference emerges: the E-I index is 0.393 for the national network and 0.331 for the international network: considering all the papers, the institutional homogeneity is almost the same in international and national sub-networks

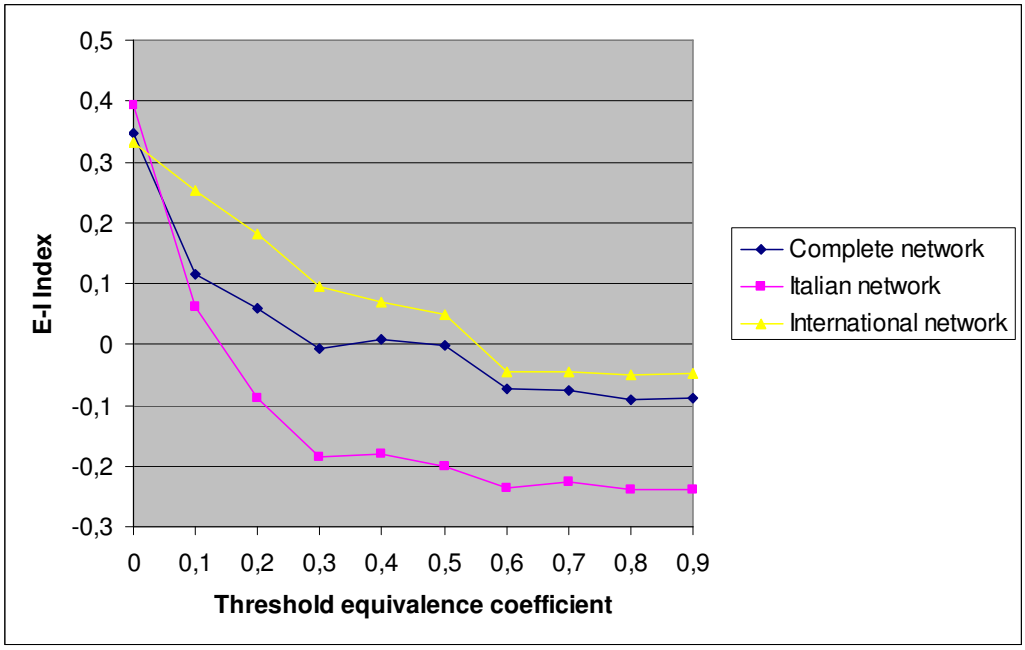
If we remove from the sub-networks the more occasional collaborations (edges with the value of of the EqC smaller than 0.1) we observe that the E-I index is significantly higher for the international network than for the national one (0.254 vs. 0.062): among the more stable collaborations, the institutional homogeneity is greater among international than national collaborations; given the above result among all the papers, it derives that, among the more occasional collaboration, the institutional homogeneity is greater among national than international collaborations. If we rise the threshold of the equivalent coefficients above 0.1, we continue to observe that the value of the E-I index is higher in the international networks; besides, as for the whole network, for both sub-networks the value of the E-I index continuously decreases (Figure 2).

So we obtain the third main conclusion of our analysis:

3. Considering all the papers, no significant difference in homogeneity exists between national and international collaborations; considering more occasional collaborations, international collaborations are more homogeneous than international one; considering more stable collaborations international collaborations are more heterogeneous than national ones.

Figure 2 shows, for the various thresholds of the equivalence coefficient (on the horizontal axis), the value of the E-I index (on the vertical axis), and this is shown for the complete network of co-autorship (blue line), the national sub-network (purple line) and the international sub-network (yellow line).

Figure 2: E-I index for different thresholds of equivalence coefficient: complete network, national and international sub-networks



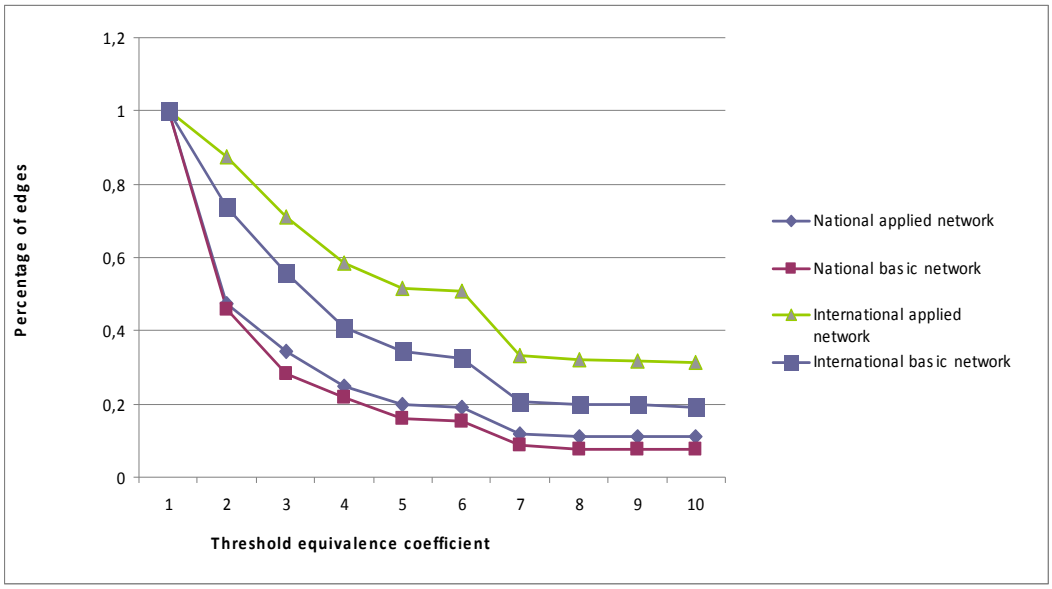
Our data let to take another dimension of analysis into consideration, that involving the content of the research, distinguished between basic and applied one. We can therefore verify if the three propositions reported above, regarding all the papers, are still valid in the two sub-groups, identified by the two kinds of research.

For this purpose, we identified four sub-groups of papers and on this basis we obtain four networks: national basic (196 nodes, 665 edges), national applied (261 nodes, 501 edges), international basic (387 nodes, 1399 edges), international applied (354 nodes, 1537 edges).

The first proposition continues to hold (there is an higher proportions of occasional collaborations in the national than in the international context), with a slight differentiation between the two kinds of research: considering the threshold of 0.1 of the EqC, 73.8% of the collaborations in the basic-international network are above this threshold, while the percentage is 45.7% in the basic-national network. The difference is even greater in applied research: the percentage of collaboration above the 0.1 threshold is 87.6% in the applied-international network and 45.7% in the applied-national network (Figure 3). Therefore, we can also conclude that the in the basic research network there is an higher ratio of occasional collaborations.

Figure 3 shows, for the various thresholds of the equivalence coefficient (on the horizontal axis), the percentage of edges that are above that threshold (the vertical axis), and this is shown for four sub-networks: national basic (purple line), national applied (blue line with rhombi), international basic (blue line with squares), international applied (green line).

Figure 3- Percentage of edges above thresholds of equivalence coefficients: national basic, national applied, international basic and international applied sub-networks



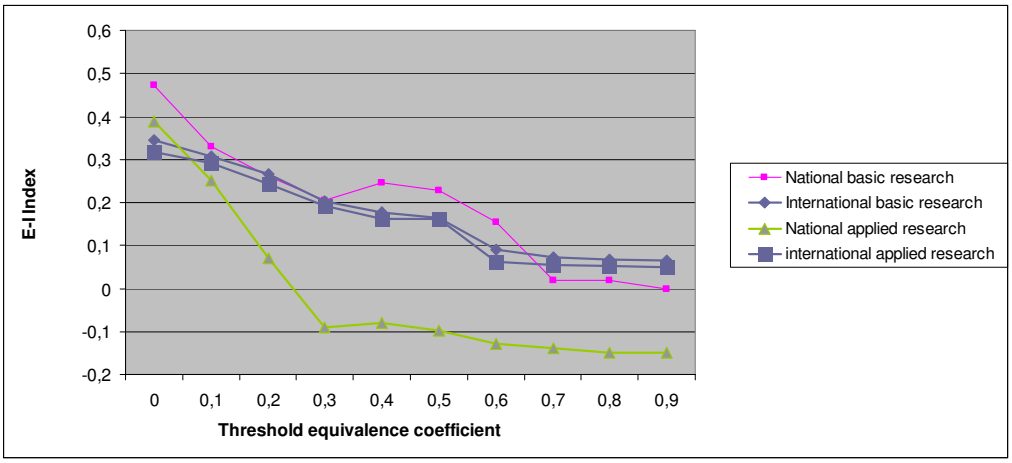
The second proposition holds for each of the four sub-networks (when the equivalent coefficient increases, the E-I index constantly decreases: the stronger ties are more homogeneous from the institutional point of view).

The third proposition is valid for the applied research: if we consider all the collaborations, they are more homogenous in international ambit, but, when we consider values of the equivalence index higher than 0.1, the E-I index of the national network strongly decreases, going below the value of the international network: when we consider the more stable collaboration, the national collaborations are more homogeneous than the international ones. For the networks of basic research, the relationship between intensity of collaboration and homogeneity is not clear. For ties with the thresholds from 0.1 to 0.3 no significant difference is found between national and international collaborations. Considering the stronger relations (EqC from 0.4 and 0.7) the international collaborations are more homogeneous from the institutional point of view, the opposite holds for the strongest relations (EqC equal to 0.8 and 0.9).

Figure 4 demonstrates what stated above: it shows, for the various thresholds of the equivalence coefficient (on the horizontal axis), the value of the E-I index (the vertical axis), and this is shown for four sub-networks: national basic (purple line), national applied (blue line with rhombi), international basic (blue line with squares), international applied (green line).

We can conclude that the distinction between basic and applied research substantially confirms the general conclusions of the previous analysis, with two further considerations: the ratio of stable collaborations in an international ambit is particularly high in the applied research; the fact that the more stable collaborations are more homogeneous in a national ambit is clearly true in the applied research, less evident in the basic research.

Figure 4: E-I index for different thresholds of equivalence coefficient: national basic, national applied, international basic and international applied sub-networks



7.1 Discussion of the results

If we look at the empirical results in the light of the third section, we find that they are consistent with the idea that, when the scientific collaboration imply different kinds of distance, different kinds of costs prevail: collaborating with spatially distant partners implies higher "costs of entry" but, once the barrier is overcome, the "cost" of each collaboration (the "marginal cost") is not particularly high, however not much higher than the costs deriving from the collaboration with close partners; on the other site, the marginal benefits deriving from each collaboration with an international partner may be assumed as higher than with a national one, as competencies may be found in wider context; this may induce the agents to collaborate frequently and this may be the reason why the ratio of stable collaborations is higher with international partners than with national one.

On the contrary, the results are consistent with the hypothesis that each collaboration with an institutionally different partner implies not only an higher "fixed cost", but even an higher "marginal cost" than a collaboration with a similar partner and this difference is not fully compensated by higher benefits; therefore the more stable collaborations happens with similar partners.

The third result of our analysis gives a new empirical contribution to the debate about the relationship between spatial and institutional distance, clarifying that a trade-off may be identified only in occasional collaborations, while a relation of complimentarity (more spatial distance, more institutional heterogeneity) holds when the relations are stable. More occasional collaborations likely reflect less ambitious projects, where the attention to the cost is more important for the agents; therefore, when the cost deriving from one kind of distance increases, the other must decrease; on the contrary, more stable collaborations rise for more ambitious projects, where the importance of specific competencies, that must be found wherever they are, is much more important than the attention to the costs.

Our analysis shows that these considerations, if correct, are particularly true for the applied research, where the illustrated effect are more evident than in the basic research-based papers.

8. 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 strength of ties and the kind of research.

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

We start from the hypothesis that both kinds of distance imply a cost, but the collaboration with a distant, on a spatial and/or institutional point of view, partner, may imply a benefit too, as the competencies owned by the distant partner could not be owned by a close one. In order to have a deep understanding of this behaviour we distinguish the collaborations according to their degree of co-autorship frequency, supposing that the choices of the institutions may be different for occasional and more frequent collaborations. Another aspect, in our opinion necessary for the analysis of scientific collaboration, is the nature of knowledge involved in the common research, if basic or applied.

We find a higher ratio of stable partnerships among spatially distant (outside the national borders) and institutionally close (of the same type) partners. We can interpret this result in the sense that the cost deriving from each collaboration involving institutional distance are particularly high, more than the costs deriving from each spatially distant partners. Then, we find that, among more stable partnerships, the two kinds of distance move in the same direction (international collaboration are more heterogeneous than national ones). Finally, we find that these results are stronger in applied research.

The contribution that this paper aims to give to the existing literature is under the profile of contents and methodology. We propose a systematic and multifaceted analysis of the complex relations between the kinds of distance, using instruments taken from the social network analysis but usually applied in different contexts, like the equivalence index, as a measure the "strength" of the relationship, and the E-I index, as a measure of homophily of linkages, therefore of institutional distance.

We desire to give, in this way, 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

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**, 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), 590–614.

Bonaccorsi A., Thoma G. (2007): "Institutional Complementarity and Inventive Performance in Nanotechnology", *Research Policy*, **36**(6), 251-283.

Broström A. (2010): "Working with Distant Researchers-Distance and Content in University-Industry Interaction", *Research Policy* **39**(10), 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), 81-100.

D'Amore R., Iorio R., Stawinoga A. (2012): "Who and Where are the Co-authors? The Relationship between Institutional and Geographic Distance in Scientific Publications", Working Paper DISES Università di Salerno, 3.321.

Dasgupta P., David P. (1994): "Toward a New Economics of Science", *Research Policy*, **23**(5), 487-521.

Gay B., Dousset B. (2005), "Innovation and Network Structure Dynamics: Study of the Alliance Network of a Major Sector of the Biotechnology Industry", *Research Policy*, **34**(10), 1457-75.

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), 57-73.

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), 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), 75-100.

Katz J. (1994): "Geographical Proximity and Scientific Collaboration", *Scientometrics*, **31**(1), 31-43.

Krackhardt D., Stern N. (1988): "Informal Networks and Organizational Crisis: An Experimental Simulation", *Social Psychology Quarterly*, **51**(2), 123-140.

Lewison, G., Paraje, G. (2004): "The Classification of Biomedical Journals by Research Levels, *Scientometrics,* **60**(2), 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), 287-316.

McKelvey M., Alm H., Riccaboni M. (2003): "Does Co-location Matter for Formal Knowledge Collaboration in the Swedish Biotechnological-Pharmaceutical Sector?". *Research Policy* **32**(3), 483-501.

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

Murray F., Stern S. (2007): "Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An Empirical Test of the Anti-Commons Hypothesis" *Journal of Economic Behaviour and Organization*, **63**(4), 648-687.

Narin, F., Hamilton, K.S.(1996): "Bibliometric Performance Measure", *Scientometrics,* **36**(3), 293-310.

Newman, M.E.J. (2004). "Coauthorship Networks and Patters of Scientific Collaboration", *Proceedings of the National Academy of Sciences,*. 101, 5200–5205.

OECD (1996): *The Knowledge-Based Economy*, Paris.

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, 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), 423-443.

Powell W. (1996): "Inter-Organizational Collaboration in the Biotechnology Industry", *Journal of institutional and theoretical economics*, **120**(1), 197-215

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

Scherngell T., Barber M. (2011): "Distinct Spatial Characteristics of Industrial and Public Research Collaborations: Evidence from the Fifth EU Framework Programme", *Annals of Regional Science* **46**, 247-266.

Van Cutsem B. (Ed.) (1994). *Classification and Dissimilarity Analysis* (LNS 93). Springer.

Wagner C.S., Leydersdorff L. (2005), "Network Structure, Self-organization and the Growth of International Collaboration in Science, *Research Policy*, **34**(10), 1608-1618.

Appendix

The following figures (Figure 1 and Figure 2) show the graphs of the national and international network of co-autorships. Different shapes represent the different kinds of institutions (while in the previous analysis we considered four categories of institutions, in the graphs below there are five categories, because the Italian biotech firms are distinguished from the other firms), different colours represent different localizations (in the graphs there is the distinction between Italian, European and Extra-European institutions, but in all the analysis we only considered the dichotomization Italy-abroad).

Figure 1. National network

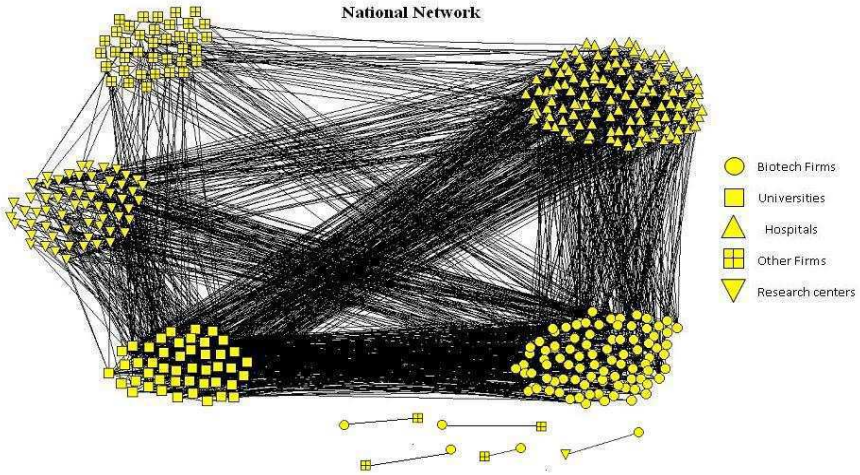


Figure 2. International network