Network Effects in R&D Partnership Evidence from the European Collaborations in Micro and Nanotechnologies

Corinne Autant-Bernard, Pascal Billand, Christophe Bravard, Nadine Massard

Network Effects in R&D Partnership Evidence from the European Collaborations in Micro and Nanotechnologies *

C. Autant-Bernard, (CREUSET, Univ. J. Monnet Saint-Etienne, France)

P. Billand, (CREUSET, Univ. J. Monnet Saint-Etienne, France)

C. Bravard , (CREUSET, Univ. J. Monnet Saint-Etienne, France)

N. Massard (CREUSET, Univ. J. Monnet Saint-Etienne, France)

Worshop DIME – Jena – 22-24 March 2007

Abstract:

Based on the research projects submitted to the 6th Framework Program of the European Union, this paper studies cooperative networks in micro and nanotechnologies. Our objective is twofold. First, using the statistical tools of the social network analysis, we characterise the structure of the R&D collaborations established between firms. Second, we investigate the determinants of this structure, by analysing the individual choices of cooperation. A binary choice model is used to put forward the existence of network effects alongside other microeconomic determinants of cooperation.

Our findings suggest that network effects are present, so that probability of collaboration is influenced by each individual's position within the network. It seems that social distance matters more than geographical distance. We also provide some evidence that similar firms (in terms of research potential) are more likely to collaborate together.

JEL classification: 031, L2, O31

Keywords: Network formation, R&D collaboration, Knowledge externalities, nanotechnologies

* We would like to thank David Frachisse and Charlène Laurent for their precious help with the data. We are also very grateful to the participants to the ADRES conference and to the ADIS seminar for their helpful comments and suggestions.

1. Introduction

Nowadays, knowledge transfers through R&D cooperation are a key research subject in economics as well as a major priority in the science and technology policy agenda of a number of countries. Recent papers in economics of science and innovation highlight the interactive and collective properties of innovation hence demonstrate the role of networks on knowledge transfers and innovation processes. Interactions among actors within scientific and technological networks of cooperation have been growing stronger over the past two or three decades – particularly in the fields of high technology. Studies so far have focused on the relations between science and technology, on inter-firm co-operations or on the collective output of research activities (such as co-publications or co-patenting). On the whole, it is shown that the capacity to produce new knowledge depends on the quality of knowledge transfers within networks, and this is considered as an output of network structural configuration (Meaghers and Rogers, 2004; Cassi and Zirulia (2004). The "small world hypothesis" often emerges from this literature linking network structure and innovation. Small Worlds (i.e. networks with short average path length and high degree of clustering: Watts and Strogatz, 1998) appears to be the structure that best characterises the empirical property of the large networks under observation[1]. By the same token, it has been theoretically shown that the rate of knowledge diffusion is highest in networks that exhibit small world properties.

Until now such a literature has rarely been articulated with the literature on the geography of innovation which tries to explain the high spatial concentration of innovative activities[2]. Indeed, on the one hand, the geographical dimension is not integrated into the analysis of networks of R&D cooperation and, on the other hand, because of the focus on the role of knowledge externalities as free and unintentional phenomena, the geography of innovation neglects the processes of cooperation between individuals. The diffusion of knowledge however, implies real links, effective interactions among actors who are embedded in networks. Therefore, one of the main results of the literature on the geography of innovation is to state that the geographical proximity as such can no longer be considered as the main determinant of agglomeration effects. It could only be the outcome of the existence of other underlying phenomena, the local networks relations as one. Breschi and Lissoni (2006) and Singh (2005), for instance, drawing their inspiration from the Social Network Analysis[3], find that social proximity is more relevant than geographical proximity when evaluating the degree of knowledge spillovers. So, when it comes to analyse the spatial characteristics of knowledge diffusion, this leads us to switch from an analysis of knowledge externalities to an analysis of accessibility to

---

[1] see Cowan and Jonard, 2004, for a review and Fleming et al. (2004) as a first empirical attempt to test for the influence of small world properties of regional networks on innovative performance.
[2] One can quote here the numerous studies in the field of the geography of innovation which have attempted to measure the geographical dimension of knowledge externalities (Jaffe et al, 1993 ; Audretsch and Feldman, 1994 are among the formers)
[3] See Granovetter (1973, 1983), Wasserman and Faust (1994).

knowledge through networks. In this case, R&D collaborations are the underlying phenomena to understand, since they determine the structure of the networks formed and are the outcome of deliberate decisions by actors.

This paper analyses the R&D collaboration networks as they emerge from the R&D projects submitted to the 6th Framework Program in the field of micro and nanotechnology. The Framework programs are one of the main instruments of the European Research Area. The aim is to promote innovation through scientific collaborations in the high value added fields of research. Such data is particularly suitable for us as it concerns R&D research projects which are the largely dominant types of arrangement for conscious production and exchange of technological knowledge between actors (Hagedoorn, 2002).

Our objective is twofold. First, following Hagedoorn (2002) on the pharmaceutical industry and Cloodt and al. (2007) on the computer industry, we characterise the structure of the R&D collaboration established between firms in the micro-nanotechnologies field. For that, we essentially use the statistical tools of the social network analysis and add a few measures of the spatial structure.

Second, we investigate what determine this structure, by analysing the individual choices of cooperation. Recent papers on the theory of network formation propose different ways of modelling the incentives schemes underlying the emerging network structures. Among these papers, those which present a collaborative conception of the relations, that is to say relations based on mutual agreements, are particularly relevant to study scientific and research networks.

Until now, the literature on this topic has been essentially theoretical. It follows the seminal work of Jackson and Wolinski (1996) regarding the determinants of cooperation and how they impact the structure of networks (Goyal et alii, 2006; Carayol and Roux, 2005). Decision to collaborate is made after considering the costs and benefits associated with the collaboration (Bala & Goyal 2000). Few authors propose empirical studies based on this theoretical background. The first attempts are notably due to Goyal and alii (2006) who reveal the formation of small words in economics using data on co-authorship. Also using data on co-authorships among economists, Van der Leij, Goyal and Fafchamps (2006) test the existence of network effects relative to other "rational" determinants influencing the decision to cooperate[4]. In this context, "network effects" refer to effects analysed in social network analysis such as those due to prior acquaintance or preferential attachment.

Similarly to these authors, we test a logit model of cooperation choice with the objective of putting forward the existence of network effects alongside other microeconomic determinants of cooperation.

[4] Some recent developments consider not only the decision to cooperate or not, but also allow the endogeneisation of the strength of the links. Such is the case, from a theoretical point of view, in Bloch and Dutta (2005), where the cost of cooperation is no longer considered as a fixed cost but is modelled as an opportunity cost of investments. From an empirical point of view, a recent paper from Goyal and Van der Leij (2006) tests the celebrated `Strength of weak ties' theory of Granovetter (1973). In a data set of collaborating economists they find support for the hypothesis of transitivity of strong ties, whilst they reject the Granovetters's hypothesis that weak ties reduce distance more than strong ties do. These studies do not integrate the geographical dimension.

Our approach however, differs from theirs in four ways. First of all, the actors are firms and not individual researchers. Second, our analysis does not concern the outputs of scientific collaboration but rather the steps leading up to the firms' decision to cooperate in R&D sponsored research projects. Third, the field analysed is micro and nanotechnology. This field is representative of a new "high tech" emerging industry for which it is particularly worthwhile using a network analysis as their borders remain somewhat imprecise. Finally, among the incentives to cooperate, we also try to estimate the role played by the geographical determinants relatively to other determinants.

This paper proceeds as follows. The data and methodological issues related to the use of this data are described in section 2. The statistical analysis of the network observed is presented in section 3. We concentrate thereafter on the explanation of cooperative links between the actors. We define the main conceptual bases of our analysis in section 4 whilst the econometric model is put forward in section 5 together with the results obtained. The conclusion in section 6 summarises the main findings and proposes further developments to improve our analysis.

2. Data description

2.1. The 6th Framework Program data base[5]

The data comes from a database made by the European division of ANRT (The French National Association for Research & Technology[6]) on the basis of information gathered by the European Commission. We examine the R&D projects submitted for the 6th Framework Program.

Each Framework Program (FP) follows a simple scheme based on strategic objectives and focusing on thematic priorities. With a budget of 17.5 billion euros for the years 2002 – 2006, the FP6 represents about 4 to 5 percent of the overall expenditure on Research and Technological Development in European Union. Projects are selected in a competitive way based on Calls for Proposals**.** Potential participants can enter into partnership with other entities to elaborate and submit a proposal to the Commission. Proposals matching the objectives and criteria given by the European Commission will be co-financed by the European Union.

Among the thematic priorities we are interested in those concerning micro and nanotechnologies: on the one hand, because it involves a lot of industrial partners and not only research centres as it is the case in some other thematic; on the other hand, because it concerns an emerging sector where the network analysis is particularly interesting. Actually, based on the identification of the actors and of their relationship, the network analysis allows us to go beyond the usual sectoral classifications to describe emerging industries. Moreover, given the significant implications of networks on the future

[5] http://ec.europa.eu/research/fp6/pdf/fp6-in-brief_en.pdf

[6] http://www.anrt.asso.fr/index.jsp

development of markets in such fields, an analysis is required concerning the way firms choose to interact when reacting to the incentives of the European Commission R&D Framework Program.

The term micro and nanotechnologies however, includes a large range of research fields ranging from nanosciences (fundamental researches) to new materials (closer to applied research) via semi-conductors. It covers, therefore, two thematic priorities: Information Society Technologies, Nanotechnologies and Nanosciences. The data we use corresponds to three different Calls for proposal concerning essentially the micro and nano systems (electronic, semi-conductors). Among the different instruments used by the Commission, we retain the lntegrated Project (IP), the Networks of Excellence and the Specific Targeted Research or Innovation Projects (STP). Such other instruments as the specific support actions (SSA) have not been considered because they do not correspond to real scientific collaborations. Finally, we must also point out that we study the network created, not only by the projects eventually retained, but by the whole propositions applied to the Commission. This is coherent with our will to consider the first steps of the decision to cooperate.

This database provides an interesting measure of “acquaintance”. Indeed, one difficulty when studying collaboration networks, and more generally social networks, come from the definition of connectedness. What is considered to be an “acquaintance” can differ considerably from one person to another and often depends on the availability of data. In fact, the data requirements constitute a problem for empirical application (Cantner and Graf, 2006). Consequently, the empirical studies analysing scientific and research networks, both in information science and, more recently, in the economics of network formation, make use of publications or patent data (either by building citation or co-authorship networks). Publications, giving information on co-authorship or co-citations, allow us to study scientific collaborations. However, this accounts mainly for public research (95% of scientific publications are produced by public institutions), and literature concerning the industrial scientific collaborations remain poor. Studies based on patent documents address this issue more directly. They look at co-patenting or patent citations, that is to say connections between inventors established via the citation of their works in the patent documents. Few scientific cooperations however, lead to a patent. Since we are interested in the connections between actors in the process of innovation, the output in terms of patent quality is not of critical importance. In this paper, using the project submitted to the European Commission, a precise definition of acquaintance is possible. Two actors are considered to be connected if they have submitted a research project to the European Commission together.

2.2. Construction of the sample

The original database contains 290 research projects. Identifying the true number of distinct entities involved in these projects, is complicated for two reasons. Firstly, two entities may have the same name. This is especially the case with the French CNRS research centres. In that case, the name has been linked with the location, entities being considered as different when the city differs. Secondly, entities may identify themselves in different ways on different projects. The European Division of ANRT has endeavoured to limit this bias by attributing a unique short name whatever the name recorded for the same entity. Even though some errors introduced by these effects may remain, we can consider that most entities are correctly identified[7].

Since our theoretical hypotheses are based on profit-seeking agents, and aim to describe how firms react to the public incentive in creating industrial cooperation, we have selected only agents involved in industrial activities (as opposed to research centres and High Education institutions). As a matter of fact, choices made by specific R&D or High education institutions may be driven by other determinants, such as objectives of the national technological policy for instance. Moreover, for econometric purposes[8] and in order to reduce the influence of punctual participations, we decided to consider only the industrial partners involved in more that one project.

The sample concerning multi-projects industrial firms contains 500 participations in 201 projects. It includes 139 entities, among which 63 SMEs. This selection of multi-projects firms does not imply important changes in the characteristics of the firms, especially in their geographical pattern. This leads however to a better balance between SMEs and non SMEs actors: whilst 62% of the total industrial entities are SMEs, SMEs represent only 45% of the multiproject sample.

The limitation of the sample to multi-project actors contributes to the reduction of the impact of marginal and punctual participations, the incentives of which can be difficult to figure out. In particular one can imagine that the criteria attached to the policy instruments may sometimes lead to the involvement of “alibi partners”[9]. Our sample could not be totally relieved of this problem. So we checked for the presence of actors who are involved in a large number of projects without ever being a coordinator or representing a significant part of the budget. None of our 139 firms enter this category.

2.3. Spatial distribution

Table 1 shows the geographical distribution of the firms and of the number of participations in projects within the sample of multi-project firms.

[7] This is a very well-known problem in the bibliometric field and different techniques of Name Matching have been proposed. Cf. for instance the proposition from M. Trajtenberg : "The names game" url = http://siepr.stanford.edu/programs/SST_Seminars/Seminar_Stanford_1.pdf

[8] The logit estimation requires reducing the number of “0” in the matrix of bilateral cooperations.

[9] For instance criteria imposing the presence of specific partners (SMEs) or imposing that the partners come from at least three different countries.

Table 1: Spatial distribution of the FP6 industrial activities in the micro-nanotechnologies (country level)

Country	firms %	Participations %
Austria	2.8	2.4
Belgium	4.3	3.6
Switzerland	4.3	8
Germany	21.6	23.2
Denmark	3.6	2.2
Spain	5.7	4
Finland	2.15	2
France	20.9	17.6
Greece	2.8	2.4
Ireland	2.15	1.4
Italy	8.6	13
Netherlands	4.3	7.4
United-Kingdom	8.6	6.8
Sweden	3.6	2.4
Other*	4.6	3.6
Core	75.4	82

* others : Cezs Republic, Hungary, Israel, Norway and Romania

Regarding the geographical dimension, we observe a high concentration of the participation (53.8 %) around three countries (Germany, France and Italy). The role of Italy is lower and closer to the one of United Kingdom in terms of number of firms (see table 1). If we compare to the characteristic of the whole industrial network (including also mono-project firms) we do not observe any differences concerning the group of dominant countries (the core of Europe: Austria, Switzerland, Germany, France, Italy, UK, Netherlands, and Belgium represents 70.6% of the whole industrial partners). Globally 79% of the firms from the core are mono-project against 83.7 % for the peripheral firms. Some countries are slightly favoured by the restriction to multiproject-partners: Switzerland, France, the Netherlands and especially Italy. By contrast, the participation of the United Kingdom and Sweden diminishes when only the multi-project partners are considered. The most striking fact is the reduction of the number of countries involved: from 32 to 19 when we consider only the multi-project partners.
Information at the NUTS[10] levels reveals that 166 nuts2[11] regions are involved in the total industrial network whereas 65 remain in the multi-project network. But the fifteen more involved regions are the same within both samples. Their share in the multi-project data basis is given in table 2.

[10] NUTS: Nomenclature of Statistical Territorial Units
[11] - NUTS 1: between 3 and 7 million inhabitants
- NUTS 2: between 800'000 and 3 million inhabitants
- NUTS 3: between 150'000 and 800'000 inhabitants

Table 2: Spatial distribution of the industrial multi-project participations in the micro-nanotechnologies (NUTS2 level)

Regions	%
OBERBAYERN	10.15
ILE DE FRANCE	10.15
LOMBARDIA	8.03
Esp. MITTELLAND	6.77
NOORD-BRABANT	5.50
PIEMONTE	2.75
MADRID	2.54
RHONE-ALPES	2.54
ATTIKI	2.11
UK55	2.11
OVERIJSSEL	1.90
DANMARK	1.90
KÖLN	1.90
STUTTGART	1,90
VLAAMS-BRABANT	1,90

3. Network description

In this section, our objective is to investigate the structural properties of the networks of partnership in micro and nanotechnologies projects of the European Program. We use the software tool Ucinet in order to analyse and to visualise the network. The methodology used in this descriptive section rests on a number of basis concepts from network analysis, which will be define first (3.1.). We then analyse the network observed by using aggregate statistics (3.2.) and individual statistics on firms (3.3.). The spatial organisation of the network is studied in the last section (3.4).

3.1. Notations and definitions

We denote by $S = \{1,2,\ldots,s\}$ the set of firms which are connected within the network. We suppose that two firms set a link with each other if both accept to participate in a common project. Then, firms are the nodes in the graph and links indicate bilateral relationships between the firms. We say that $g_{i,j} = 1$ (and equivalently $g_{j,i} = 1$) if firms *i* and *j* have set a link with each other, otherwise $g_{i,j} = 0$. Thus, the information on projects and participating firms allows us to construct a network of collaboration. We denote by *g* the network formed by the participating firms. Let *L* be the number of links in the network and L_i be the number of links formed by the firm *i*. The average density of a network, denoted by β, is a measure of the connectedness of the network. β is the ratio of the number of links observed over the maximum number of links that we can have in the network. In a completely connected network, we have $\beta = 1$. We call the degree of a firm the number of links of this firm. The average link degree is the average number of degrees of participating firms.

We say that there is a chain between firms i and j in a network g, either if $g_{j,i} = 1$ or there exists a set of distinct firms $k_1, k_2, \ldots, k_n$, such that $g_{i,k1} = g_{k1,k2} = g_{k,ni} = 1$. We call a (connected) component of a network g a non-empty subnetwork $g' \subset g$, such that: if i and j belong to this component, then there exists a path between these two firms, and if i belongs to this component and j does not belong to it, then a path between i and j does not exist. We denote by C the number of components in the network.

We can order the components in terms of their size and we will say that a network has a giant component if the largest component fills a relatively large part of the graph and all other components are small (Goyal, et al., 2004, p. 6).

We call the geodesic distance between two firms i and j in a networks g the length of the shortest path between these firms in g. Let $d(i,j;g)$ denote the (geodesic) distance between i and j in g. If i and j are not connected, we set $d(i,j;g) = \infty$. In case g is connected, the average distance between nodes in a network g is given by

$$d(g) = \frac{\sum_{i \in S} \sum_{j \in S} d(i,j;g)}{s(s-1)}$$

Where g is not connected, we shall use the average distance in the components (or in the giant component) as a proxy for the average distance within the network.

We call the clustering coefficient of a network g a measure of the correlation between links of different individuals. More precisely, the clustering coefficient of a firm i is given by

$$C_i(g) = \frac{\sum_{l \in L_i} \sum_{k \in L_i} g_{l,k}}{l_i(l_i - 1)}$$

This ratio tells us what percentage of a firm's collaborators cooperates with each other. We call the clustering coefficient of a network g the average of the clustering coefficients over all firms in g. The clustering coefficient of a network measures the transitivity of the network.

We define a star network as a network where a firm i has set up a link with all the other firms whilst they have set up one link only, a link with the firm i. In that case, firm i is called the centre of the star. Note that in a star network, we have $C_i(g) = 0$.

Following Goyal and al. (2004), we can say that a network g exhibits small world properties if it satisfies the following properties:

1. The number of nodes is very large as compared to the average number of links.
2. The network is integrated; the large component exists and covers a large part of the population.
3. Clustering is high.
4. The average distance between nodes in the giant component of size s is small, $d(g)$ is of the order $ln\ (s)$.

3.2. Aggregate statistics

The main aggregate statistics are presented in Table 3 and Table 4 concerning the network observed.

Table 3 tells us that firms participating in more than one project are present in less than four projects on average, which is relatively few. This result partly explains why the average density of the network is so low (*β = 5 %)*. Such a low level of connectedness among firms may influence knowledge diffusion. Indeed, information can flow easier or faster in a highly connected network, allowing important discoveries and scientific information to reach rapidly most members of the network.

Table 3: Network statistics

Indicator	Network of Multiproject partners
Number of projects	201
Number of entities (S)	139
Mean project per firms **Standard deviation** **Minimum** **Maximum**	3.583 3.985 2 30
Mean entities per project **Standard deviation** **Minimum** **Maximum**	2.478 1.588 1 9
Number de components (C)	7
Number de links (L) **Number of potential links (L_{max}=S(S-1)/2)** **Average Density : β (L/ L_{max})** **Standard deviation of density**	482 9 591 0.0503 0.2185
Average link degree **Standard deviation link degree** **Minimum link degree** **Maximun link degree**	6.935 6.843 0 43
Network centrality (betweeness)	19.82 %
Average relative Betweenness **Standard deviation** **Min** **Max**	1.103 3.046 0 20.783
Clustering coefficient **Weighted Clustering coefficient**	0.599 0.316

We now turn to the number of components within the network. We observe that this number is high relative to the total number of firms (7 components for 139 firms). However, table 4 tells us that the level of integration in the network is very high, since the largest component contains 94.2% of the total number of firms (131 firms over 139). The other components are small. They contain at best 2 firms[12]. Thus, we can say that the network consists of a giant component and of a number of isolated groups of firms which can not be the vectors of the diffusion of knowledge across projects.

Table 4: Analysis of the main component

Size (number of firms)	131
Number of links (L)	480
Number of potential links (L_{max})	8515
Density : β (L/L_{max})	0.0564
Average distance **Maximum distance**	2.70182 6

The transitivity of the network is measured by the clustering coefficient within the network. In order to characterise the level of clustering, we observe from Table 3 that each of the 139 firms has 5.1 links on average. In other words, the probability for a link to be formed is approximately 0.03. In a random graph, since the probability of link formation is independent, the clustering coefficient should be approximately equal to this probability of a link. However, the clustering coefficient is 0.56 which is very much higher. The existence of projects including more than two firms increases the clustering coefficient, so it may largely explain the result.

We now examine the (geodesic) distance between firms. We observe in Table 4 that this distance is relatively low. For instance, if we consider the giant component, the average distance is a little less than 3.

To conclude, we can say that the network exhibits small world properties since it satisfies the four following properties: large number of nodes relatively to the average number of links, existence of a large component which covers a large part of the population, high level of clustering, small average distance between nodes in the giant component, less of the order *ln (*s), that is to say 4.9.

3.3. Individual firm statistics

Figures 1 and 2 give us some indications concerning the distributions of the number of projects and the distributions of the number of links among firms in our sample. We observe that both distributions are very unequal. Indeed, concerning the number of projects in which firms are involved, most of the firms are involved in two projects only. Moreover, there exist few firms involved in a great number of projects. More precisely, 5 firms participate in more than 13 projects. In fact, 5% of the firms contribute to 27% of the participations. Concerning the degree of firms, most of the firms have less than 10 links, but few firms have more than 35 links. Four firms (3%) are linked with more than 35

[12] Some components contain only one firm because of the selection of multiproject firms. Indeed, firms that created links only with mono-project firms have no partners among the multiproject firms.

partners and represent nearly 32% of the links. By contrast, 12% of firms are involved in 2.7 % of the links.

The last result is to be expected since for a firm *i* there will be a correlation between the number of projects in which firm *i* is involved and the number of links that this firm has. Due to this heterogeneity, if firms are the main vectors of knowledge flows, few firms have a central role in the network.

Figure 1: Distribution of the number of projects within the network

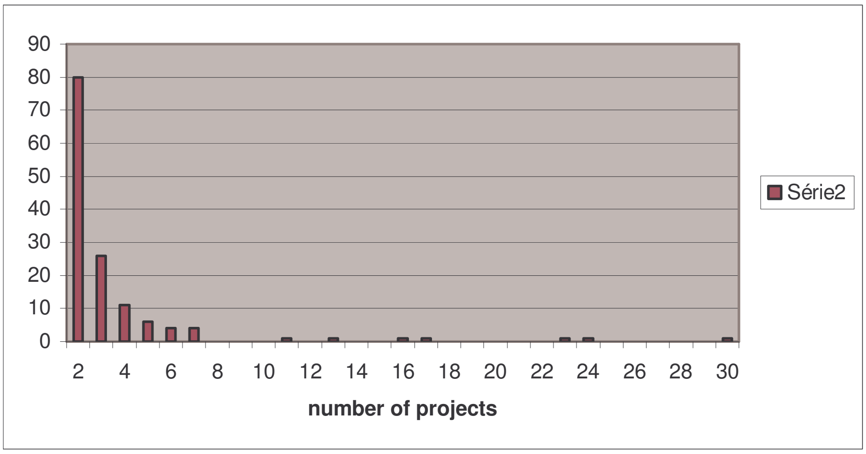
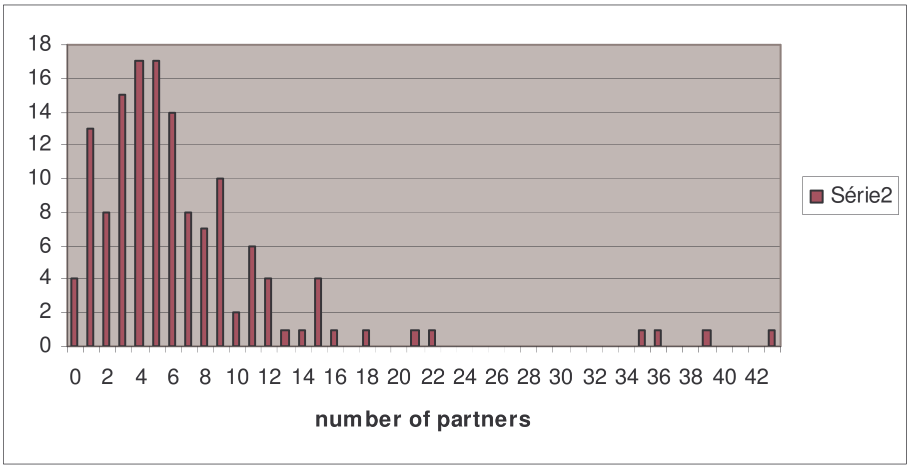


Figure 2: Distribution of the links within the network



In order to assess the role of these well connected firms, we need to know how they are linked with each other and with the other multi-project firms, within the network (note that all the well connected firms belong to the giant component). In order to do this, we have plotted the local network around the four most connected firms (firms 251, 638, 1052, 1287 in Figure 3 annexed). Figure shows that the network can be viewed as constructed around the central role of these 4 firms, forming an "inter-linked star”. Goyal and al. (2004) call a network an inter-linked star if it is divided into two groups of agents, central agents and peripheral agents. Central agents are interconnected with each others and each of them has links with peripheral agents, while each of peripheral agents has one link only, a link with one of the central agents. In an inter-linked star communication network, central agents play a key role since each one controls the transmission of information towards the peripheral firms. In the network of R&D collaborations, central firms may play a great role by choosing to diffuse or not the knowledge across firms. The network of collaboration in micro and nanotechnologies however, is not a pure inter-linked star. Indeed, there exist firms which are neither central nor peripheral agents. In particular, some peripheral agents are not exclusively linked with one of the central firms (the firms at the centre of the figure 3). Therefore, the role of firms 251, 638, 1052, 1287 in the diffusion of knowledge between projects is likely to be smaller than if the network was a pure inter-linked star. In particular the betweenness index for these firms is not very high.

Table 5: Degree of link and betweenness of the four better connected firms

firms	degree	betweenness
1287	44	0.196
1052	41	0.146
638	37	0.121
251	35	0.207

There exist different means to evaluate the role of these firms. Following the procedure used by Goyal and al. (2004), we can randomly delete 4 of the firms from the giant component and then compare the properties of the new network with those of the case where we deleted the 4 most connected firms. We can see in Table 6 that the removal of 4 firms at random has almost no effect on the main characteristics of the network whilst, on the contrary, the removal of the 4 most connected firms has a significant effect on the structure of the network. In particular, it contributes to a reduction of the size of the giant component associated to an increase of the average distance (from 2.7 to 3.38), of the maximum distance (from 6 to 7) and of the clustering coefficient (up to 0.39). These observations confirm the star position of the four most connected firms, even if their star role is weak.

Table 6: Influence of the 4 most central firms

	Total network	Network w/o top 4 firms	Network w/o random 4 firms
Size of the giant component	0.942	0.911	0.94
betweeness (network centralization index)	19.82	13.46	20.44
average distance (among reachable pairs)	2.67	3.38	2.71
weighted clustering coefficient	0.31	0.389	0,31
Maximum distance	6	7	6

Compare to previous findings on collaborative networks (such as those based on co-publications between researchers), the network of inter-firm cooperation based on the incentives of the European Framework Program display a different structural form. Few firms do play a crucial part, but the emerging form is less hierarchical than usually observed for scientific collaboration in the sense that clustering is more global and does not only concerned separately the relationship between central firms on the one hand and the relationship between more peripheral actors on the other hand.

3.4. Geographical organisation of the network

The geographical pattern of the network confirms the concentration in the European Core countries. We find that 330 out of the 482 cooperation links are internal to the core of Europe, that is to say that they involve partners from Austria, Switzerland, Germany, France, Italy, UK, Netherlands, and Belgium. On the contrary, only 5% (25) of the cooperation links involve two partners coming from the rest of Europe.

We also note that 88.5% are international cooperation links whilst 9% are national and 2.5% are regional. Since the European Union enforces at least three of the partners to come from distinct countries, this large share of international relationships is not surprising.

Another remarkable aspect concerns the location of the most central firms. The four central firms identified above are all located in the core of Europe. Moreover, among the 20 % of the most central firms (in terms of degree and betweenness), 81.5% are located in the core and 11.1 % in Scandinavia.

The next step consists in trying to understand the mechanism and individuals incentives that lead to such a structural and geographical form.

4. R&D collaborative networks: main hypothesis

In this section we try to explain the collaboration network observed. We wish to understand why firms engage in collective R&D projects and so construct collaborative networks with such structural characteristics. Consequently, our level of analysis is the micro-level of individual firms (the nodes of our network).

Firms will engage in cooperation with another firm only if they expect to gain from this collaboration. This leads us to investigate how the characteristics of each partner and of their bilateral relationship influence the decision to collaborate. Another focus of our work is on the determinants of the cooperative linkages that overcome the bilateral characteristics and reveal networks effects. Finally hypothesis concerning the role of the geographical dimension have to be assessed with the following underlying question: is the geographic concentration observed only due to the distribution of opportunities of cooperation over space or are there cumulative effects based on the role of geographic proximity?

3.1 Individual features

Two main individual features may explain the highly concentrated distribution of the cooperation's links among firms: the firm research potential and its absorptive capacity. Indeed, one can firstly put forward the hypothesis that there exists a size attractive effect, firms being more likely to choose large R&D firms as partner.

- Hypothesis 1: the expected profit of cooperation depends on the research potential of each of the individuals that compose the pair.

Such an agglomeration around firms displaying a strong research potential can also be reinforced due to the absorptive capacities constraints. Indeed, numerous studies have shown that absorptive capacity (that is to say existence of internal research potential) is a necessary condition to benefit from knowledge transfer through cooperation. The larger is the gap between two entities, the less each individual can benefit from the cooperation. Those which display the weaker potential are of minor interest to those which have a high research capacity and, at the same time, they cannot plainly exploit the potential of the stronger partner because of their inadequate absorptive capacity.

- Hypothesis 2: The gap in research capacities between the two entities is likely to reduce the probability of their cooperating.

This problematic is particularly interesting in the European context, given that one of the main objectives of the European Commission is to promote the participation of the SME to R&D cooperation.

3.2. Network effects:

These network effects are threefold. They refer to the diffusion of knowledge externalities, to the creation of opportunities through social distance as well as to reputation effects.

Network effects can appear as sources of externalities facilitating the diffusion of knowledge within the network. For instance, if there is a link between i and j, and a link between j and k, then firm i might access the resources of firm k due to the collaboration with firm j.

- Hypothesis 3: the benefit from the cooperation increases, firstly, with the number of partnerships tied to a firm and also, with the number of links formed by its own partners.

In the context of cooperation between firms this, however, cannot be disconnected from the problems of appropriability. Indeed, firms are highly conscious of the problem of control or knowledge appropriability that can emerge from the diffusion of information amongst them. So, they generally face a trade-off between the necessity to increase the number of their direct and indirect partners in order to absorb new knowledge and the necessity to reduce this number so as to control the dissemination of their own knowledge. This can explain the star and small-worlds structures of real / research networks that allow the multiplication of direct links, with partners themselves developing few links elsewhere. In this case, the effect of the number of indirect links per direct link of a firm might be negative.

- Hypothesis 4: the benefit from the cooperation decreases with the number of indirect partners per direct partners due to appropriability problems

Together with knowledge externalities, the networking structure of research may create opportunities through social distance. Indeed, previous to all decisions, network effects are also the bases of the creation of opportunities to cooperate. Between the situation of perfect information on the one hand or of random matching on the other hand, there exists an intermediary situation where information concerning possible partners diffuses through "prior acquaintance" within a network or by the mechanism of « preferential attachment » (Barabasi and Albert (1998); Newman (2001,2004). The main idea therefore, is that firms form alliances because they are embedded in social networks.

The embeddedness of actors in social networks contributes to determine the opportunities to cooperate so, in our case, this is likely to reduce the cost of a first cooperation. Even if the likelihood of collaboration between two firms depends on their characteristics, in order to initiate a collaborative research project, two firms first have to meet. So, information about potential partners must circulate among firms and many authors have shown the role of prior common acquaintances in this process of information circulation reducing the cost of meeting (see Granovetter, 1995, and Fafchamps and al., 2006 for an empirical test of these phenomena within research networks). However, this creation of opportunities is only relevant for partners who have not collaborated before. In the case of partners who know each other, one can only consider the self-reinforcement process due to the benefit of having learned to cooperate together.

- Hypothesis 5: The probability of cooperation depends on the number of common acquaintances between firms *i* and *j*: number of common partners or numbers of projects in which these two firms have participated together.

Finally, many authors show that the dynamic of scientific networks is largely explained by the mechanism of preferential attachment based on reputations and rewards. The collaboration choices of scientists are motivated by reward structures within science where co-authorships, citations and other forms of professional recognition lead to additional work and enhanced reputation in a virtuous circle. Newman (1991), Barabasi and Albert (1999) reveal the structural characteristics of scientific networks based on preferential attachment. Is such a mechanism identifiable within networks of collaborations between firms? It is important to investigate this question because, even if such path-dependence and self-organising processes are likely to lead to an increase in the connectivity within networks, they may also contribute to an increase in the degree of concentration of these networks. The objectives of public policies concerning the equity of participation and of equal distribution of knowledge through cooperation could thereby be opposed.

Hypothesis 6: The probability of cooperation depends on the degree of the pair

3.4. Geographical distance

The networking dimension of R&D partnership may be closely related to the geographical dimension. Firstly, space may affect the architecture of the network. Indeed spatial distance can increase the coordination costs and reduce knowledge diffusion. One can therefore expect a spatial concentration of partners. Some recent studies of the "geography of innovation" however, (Breschi and Lissoni, 2006 for instance) observe that what is usually considered as spatial effects would mainly rely on network effects. Testing whether some specific effects due to geographical proximity remain, once controlled for individual features and network effects, is an important issue.

- Hypothesis 7: the probability to cooperate depends on the geographical distance that separates the partners.

Secondly, the increasing interconnectedness of scientists within and across countries can conversely affect the geographical distribution of innovative activities. Are these changes likely to increase the economic gap between countries? Do they enhance a core-periphery structure where leading countries cooperate with each other, or do they encourage a symmetric configuration where lagging countries benefit from the cooperation with leading ones?

- Hypothesis 8: the probability to cooperate is higher within the core of Europe.

The empirical model estimated in the following section will attempt to address these two geographical issues.

4. Formation of collaborative networks in nanotechnologies: econometric model

The decision to engage in a partnership is estimated by means of a binary choice model. The profit associated to the formation of one link between two entities i and j is considered as the latent variable. Since a measure of this profit if not available, we use a logit model where the dependent variable takes value 1 if i and j collaborate and 0 otherwise. Collaboration is assumed to occur if the associated profits are positive.

4.1. Estimation issues

For each pair of firms, a link is assumed to be created if and only if the expected profit, noted Π_{ij} is positive:

(1) $\Pi_{ij} > 0 \quad with \quad i = 1,\ldots s, \quad j = 1,\ldots s \quad and \quad i \neq j$

so the probability that the link *ij* is formed is $P_{ij} = P(\Pi_{ij} > 0)$.

The profits of each link, assumed to give symmetrical profit to both firms *i* and *j*, is given by:

(2) $\Pi_{ij} = V_{ij} + \varepsilon_{ij}$

where V_{ij} is a function of all the characteristics of *i* and *j*. ε_{ij} is a random perturbation.

We choose a linear expression for V_{ij} :

(3) $V_{ij} = \beta X_{ij}$

where X_{ij} is the vector of the observable characteristics of *i* and *j* and β is the vector of the parameters to be estimated.

These characteristics may be associated with the firm itself (X_{ij}^{ind}) but also with its position within the network (X_{ij}^{net})

Hence the following latent model:

(4) $\Pi_{ij} = X_{ij}^{ind}\beta^{ind} + X_{ij}^{net}\beta^{net} + \varepsilon_{ij}$

This will allow us to test whether only individual characteristics affect the probability to create a link or not. The hypothesis of network effect will be tested by looking at the sign and significance of the β^{net} vector of parameters.

4.2. Construction of the variables

Equation (4) indicates that empirical measures for individual characteristics and their position within a network are required. Additionally, the pairs (*i, j*) who create a link must be identified. Two firms are considered connected if they have submitted a proposition to the European Commission in the 6th Framework Program together. This seems a reasonable definition of scientific acquaintance: most firms that have submitted a project together will know one another quite well. The dependent variable is a binary one, taking value 1 if entities *i* and *j* submitted a proposal together and 0 otherwise.

As in Fafchamps, van der Leij and Goyal (2006), the symmetrical and undirectional dimension of the network regression raises a practical difficulty. Regressors must be introduced in the regression in a symmetrical way. This requires the explanatory variables to be built so that we obtain the same regressors, whatever the order of indexation.

Following Fafchamps, van der Leij and Goyal (2006), we choose the absolute difference and mean:

(a) $\Delta x \equiv \left| x_i - x_j \right|$

(b) $\bar{x}_{ij} \equiv \dfrac{x_i + x_j}{2}$

This will give the same value if we reverse the order of *i* and *j*.

All the explanatory variables presented below are built according to expression (a) and (b), except one spatial variable that are relative to the pair (*i, j*) and not to each entity.

Table 7 gives the definition of each variable and indicates whether it is expressed as a difference, noted (a) or as a mean, noted (b).

The R&D potential of each pair (*i,j*) is accounted for by means of the number of projects submitted by each entity *i* and *j (MProjects)*. This variable allows us to control for the degree of participation of each firms but however, is an imperfect measure of the R&D capability of an entity. A more accurate indicator is built using the financial mount associated to each project in which the entity is involved. Two variables are derived from this information: *MBudget* which accounts for the average budget of the projects submitted by each entity and *MSharebud* which accounts for the share of the financial support obtained by each entity (expressed as the average percentage over all the projects in which the entity is involved).

In order to account for the technological gap between *i* and *j* and absorptive capacity of firms, these three variables are also introduced as an absolute difference between *i* and *j*. A positive sign is thus expected for the variables expressed as the mean while a negative sign is expected when their expression is absolute difference.

A third set of variables controls some individual characteristics which are likely to affect the decision to collaborate: the leading position of *i* and *j* in the projects recorded in the FP6[13], and the dummy taking the value 1 if *i* and *j* are SMEs. Other dummy variables are also introduced in order to indicate the difference within the pair: *OneLeaders* (respectively *OneSME*s) take the value 1 when only one of the firms is *leader* (respectively *SMEs*) and 0 otherwise.

[13] Each proposal submitted to the Framework Program has to mentioned which partner will act as the leader. It does not mean that this entity is the most important or the most central partner. It usually characterises the role of interface played by this entities towards the European Commission.

Table 7: Description of the explanatory variables

Hypothesis	Variable name	Variable description	Type of variable
1-Research potential	***MProject***	Number of projects submitted by *i* and *j* *	**(b)**
	MBudget	Total amount of financial support obtained on these projects *	**(b)**
	MShareBud	Share of the total amount of financial support obtained by *i* an *j*	**(b)**
2-Absorptive capability	***DProject***	Gap in the number of projects submitted by *i* and *j* *	**(a)**
	DBudget	Gap in the total amount of financial support obtained on these projects *	**(a)**
	DShareBud	Share of the total amount of financial support obtained by *i* an *j*	**(a)**
3-Other individual effects	***Leader***	Dummy variable indicating whether *i* and *j* are leading entities (1) or not (0)	
	DifLeader	Dummy variable indicating whether *i* or *j* are leading entities (1), 0 otherwise.	
	SME	Dummy variable indicating whether *i* and *j* are Small and medium enterprises (1) or not (0)	
	DifSME	Dummy variable indicating whether *i* or *j* are Small and medium enterprises (1), 0 otherwise.	
4-Network effects	***Netdist***	Inverse of the geodesic distance between *i* and *j* in the 5th FP	
	ComProj	Number of common projects to *i* and *j* in the 5th FP	**(b)**
	MPart1	Number of direct partners of *i* and *j* (distance 1)	**(b)**
	MPart2	Number of indirect partners of *i* and *j* (distance 2)	**(b)**
	MPart3	Number of indirect partners of *i* and j (distance 3)	**(b)**
	RatioPart	Ratio of the number of indirect partners on the number of direct partners	**(b)**
5-Spatial effects	***Neighbour***	Dummy variable indicating whether *i* an *j* belong to neighbouring countries or not	-
	Core	Dummy variable taking the value 1 if both *i* and *j* belong to the 8 European leading countries or both *i* and *j* belong to the peripheral countries	-

* *variables transformed in logarithm.*

Our main regressors of interest are described in part 4 of the table. Their aim is to determine whether two firms collaborate in R&D only because they have common interests, or whether their position within the network matters. Where the latter is true, the number of partners of *i* and *j* might affect the probability of collaboration. Such network effects can come from knowledge externalities. Choosing a partner can be not only a way to access to its particular competences, but also a way of benefiting from its network of collaborators and accessing indirectly the competences of its partners. The greater the number of relationships, the more indirect knowledge flows. In order to test this effect, the position of *i* and *j* within the network is measured by their number of partners. Three levels of partnership are considered, according to the number of links that separates two entities. The first level (*MPart1*) gives the number of partners at a distance 1 of *i* and *j*, that is to say all the entities involved in a common project respectively with *i* and *j*. *MPart2* and *MPart3* measure the number of indirect partners for *i* at distance 2 and 3 respectively.

$$Mpartk = \left|\{j / d(i, j; g) = k\}\right|, k \in \{1;2,3\}$$

Partners of a higher order might have been considered. Their number however, drastically increases after distance 3 and their number becomes almost the same for all entities. As shown in the next section, this very low level of dispersion already raises some difficulties when the variable *MPart3* is introduced in the model. Let it be noted here that the number of partners is observed for the whole data base so it does not only include industrial partners but all the different types of actors involved in European R&D proposals in the micro-nanotechnology field. In order to deal with the problematic of appropriability, we also constructed a variable indicating the average number of indirect partners for each direct partners of a pair (noted *RatioPart*). We anticipate that the more numerous are the indirect partners per direct partners, the more difficult is the control on the knowledge diffusion. Temporal network effects are also introduced through two other variables which describe the relational proximity of firms *i* and *j* within the network formed by the projects developed in 5th FP[14]. The first one indicates the number of common projects to the firms that has been financed within the 5th FP (*ComProj*). Indeed if firm *i* cooperates with firm *k* and firm *j* cooperates with *k*, this should increase the probability that *i* and *j* create a link. This will then allow us to construct an indicator of distance inside the network based on the minimal number of links between *i* and *j* during the previous period. Precisely, the second indicator gives the geodesic distance between firms *i* and *j* within the 5th FP network (*Netdist*). It will thus be possible to evaluate the role of “prior acquaintance” or “preferential attachment” in the formation of new partnerships.

Finally, two variables account for spatial effects. The first one measures the impact of the geographical proximity on the probability of collaboration. It takes value 1 if *i* and *j* belong to neighbouring countries (Neighbour). A positive sign of the associated coefficient would highlight that proximity matters in R&D partnership decisions. The second variable is introduced to test whether a link is more likely to be created when both entities belong to networks from the core of Europe (Austria, Switzerland, Germany, France, Italy, UK, Netherlands, and Belgium) or when both belong to

[14] 40 firms out of the 139 of our sample were already involved in the 5th FP.

networks from the periphery (CORE). A positive sign of the coefficient will indicate that the formation of collaborative network in micro-nanotechnologies reinforces the core-periphery structure, while a negative sign will give some evidence that R&D partnership may be a way for peripheral countries to benefit from the competences of those in the lead.

4.3. Results

Table 8 give the results obtained for different specification of the model. Estimation (1) and (2) introduce individual features only. Estimation (3) to (6) tested the network effects. The variable NETDIST and COMPROJ cannot be introduced simultaneously. Indeed, if COMPROJ is different from zero, then NETDIST equals 1 (Since the highest Pseudo-R^2 is obtained with COMPROJ, the remainder of the estimations focus on this specification). Similarly, the variable MPART and RatioPart are introduced alternatively. Finally, the spatial variables are added in estimations (7) and (8).

Three main results arise. First of all, the main determinants are due to individual characteristics. Both the R&D potential and the absorptive capability of *i* and *j* matter. Once controlled for the number of projects submitted by *i* and *j* (*MProject*), the average financial size (*MBudget*) increases significantly the probability of collaboration. The share of the financial support obtained by *i* and *j* on these projects also has a positive effect, in spite of a much lower coefficient. By contrast, the variables expressed as an absolute difference between *i* and *j* reduce the probability of forming a link. The more *i* and *j* differ in R&D potential, the less they collaborate. Such a negative relation supports the hypothesis of the role of absorptive capacities. Similar firms (in terms of R&D dynamism in micro and nanotechnologies) are more likely to collaborate together.

The impact of the other individual variables is significant. A negative effect is associated to leading firms and SMEs. In other words, the probability to collaborate is higher when both firms are large and non-leader. The negative effect associated to leading firms may reveal a preference for the leading role of non profit institution, due to the incitation of the European Commission or for fear of knowledge appropriability when the project is driven by a firm.

Secondly, network effects do occur. The most noticeable effect is the positive impact of the relational proximity in the 5^{th} FP, measured either by COMPROJ or by NETDIST. Hence, firms are more likely to cooperate when they have already collaborated, or when the social distance between them is small enough.

Table 8 : Logit estimation of collaborative choices

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constante	-15.93*** (2.18)	-15.72*** (2.34)	-14.88*** (2.35)	-14.60** (2.35)	-22.29* (13.00)	-8.08*** (2.76)	-7.98*** (2.76)	-21.67* (13.1)
MProject	2.11*** (0.28)	3.16*** (0.30)	3.14*** (0.30)	3.10*** (0.30)	2.06*** (0.38)	2.29*** (0.38)	2.27*** (0.38)	2.04*** (0.39)
MBudget	0.87*** (0.14)	0.80*** (0.15)	0.75*** (0.15)	0.73*** (0.15)	0.44*** (0.16)	0.44** (0.17)	0.42** (0.17)	0.43*** (0.16)
Msharebud	0.002 (0.017)	0.04** (0.01)	0.04** (0.01)	0.03* (0.01)	0.09*** (0.02)	0.07*** (0.02)	0.07*** (0.02)	0.09*** (0.02)
DProject	-0.43*** (0.11)	-0.33*** (0.12)	-0.31** (0.12)	-0.29** (0.12)	-0.20 (0.13)	-0.28** (0.12)	-0.28** (0.13)	-0.204 (0.13)
DBudget	-0.29*** (0.03)	-0.31*** (0.03)	-0.30*** (0.03)	-0.31*** (0.03)	-0.29*** (0.04)	-0.32*** (0.04)	-0.32*** (0.04)	-0.29*** (0.04)
Dsharebud	-0.01 (0.12)	0.002 (0.01)	0.005 (0.01)	-0.007 (0.01)	-0.007 (0.01)	-0.01 (0.01)	-0.009 (0.01)	-0.006 (0.01)
LEADER	-	- 5.03*** (0.55)	-5.43*** (0.68)	-5.10*** (0.57)	-5.64*** (0.70)	-5.59*** (0.70)	-5.59*** (0.70)	-5.63*** (0.70)
DIFLeader	-	- 2.43*** (0.16)	-2.42*** (0.16)	-2.42*** (0.16)	-2.40*** (0.16)	-2.44*** (0.16)	-2.43*** (0.16)	-2.39*** (0.16)
SME	-	-0.36* (0.18)	-0.38** (0.19)	-0.35* (0.19)	-0.25 (0.19)	-0.38** (0.19)	-0.40** (0.19)	-0.27 (0.19)
DIFSME	-	-0.42*** (0.11)	-0.41*** (0.11)	-0.40*** (0.11)	-0.33*** (0.11)	-0.40*** (0.11)	-0.41*** (0.11)	-0.34*** (0.11)
COMPROJ	-	-	1.47*** (0.42)	-	1.36*** (0.43)	1.42*** (0.43)	1.40*** (0.43)	1.34*** (0.43)
Netdist	-	-	-	1.10*** (0.33)	-	-	-	-
MPart1	-	-	-	-	0.013*** (0.004)	0.012*** (0.004)	0.012*** (0.004)	0.013*** (0.004)
MPart2	-	-	-	-	0.001*** (0.0006)	-	-	0.001*** (0.0006)
MPart3	-	-	-	-	0.007 (0.009)	-	-	0.007 (0.009)
RatioPart	-	-	-	-	-	-0.008*** (0.002)	-0.008*** (0.002)	-
Neighbour	-	-	-	-	-	-	0.002 (0.01)	0.001 (0.01)
CORE	-	-	-	-	-	-	0.202* (0.117)	0.156 (0.118)
Pseudo-R²	0.151	0.262	0.266	0.264	0.276	0.273	0.273	0.277
AIC	0.339	0.296	0.294	0.295	0.291	0.292	0.292	0.291
LR-test	579.18***	1002.20***	1017.64***	1012.58***	1057.18***	1044.02***	1047.12***	1059.01***
Obs.	9591	9591	9591	9591	9591	9591	9591	9591

The firms' position within network also appears as a determinant of decision to form a partnership. The higher the number of direct partners (*MPart1*) of *i* and *j*, the higher the probability of collaboration. One might guess that the positive sign of the coefficient associated to MPart1 captures mainly network effects. Disentangling these effects from a purely statistical effect however, remains problematic. Indeed, the probability of *i* and *j* collaborating will mechanically depend on their total number of links. The probability of *i* choosing *j* as a partner increases with the number of projects submitted by *i* and *j*, as well as with the number of partners they have. For this reason, we must be cautious when interpreting this variable. Two elements are however, likely to reduce the mechanical characteristics of this relation. Firstly, the pure size effect due to the potential of the pair is already controlled by way of the R&D potential variables. Secondly, the number of partners observed include all types of partners and not only industrial ones.

The effect produced by the number of indirect partners seems less important. At a distance 2, the number of partners has a positive effect on the probability of collaboration. But the coefficient is very weak. The coefficient of the number of partners at a distance 3 (*MPart3*) can hardly be interpreted. The very low level of dispersion of this variable (see appendix) prevents a correct estimation of this effect. To better deal with the concomitant effects of positive network externalities on the one hand and of the risk of uncontrolled knowledge diffusion, on the other hand, we introduce a variable indicating the number of indirect partners per direct partners (*RatioPart*). As anticipated, it has a significantly negative effect. This means that if firms seek to benefit from knowledge externalities through their partners, they also try to limit the number of partners of their own partners in order to reduce the risk of uncontrolled knowledge diffusion.

Finally, the spatial dimension does not seem to play a part in the partnership decisions. Spatial distance (accounted for by the variable Neighbour) is not significant. Entities located in countries sharing common borders do not collaborate more frequently than distant firms. Moreover, there is not any effect revealed by the variable CORE. The strong concentration of the nodes and links of the network in the core countries (list…) does not mean that the choice of a research partner relies on the geographical proximity, once controlled for the other factors. Then, the spatial concentration would be due to the location of opportunities and no self-reinforcing effect would occur between Core countries.

5. Conclusion

Based on the research projects submitted to the European Commission, R&D partnership in micro and nanotechnologies exhibits several specific features. First, the network resulting from firms' collaboration satisfies most of the small world properties, with short average path length and high degree of clustering. Second, as already observed for other scientific networks, few actors play a crucial part. However, the network structure is less hierarchical than observed in other networks, some firms playing an intermediate role between central and peripheral firms. Third, the geographical pattern is characterized by a strong concentration of firms and links in the European Core Countries (Austria, Switzerland, Germany, France, Italy, UK, Netherlands, and Belgium).

However, a simple description of the network does not allow us to understand the mechanisms and individual incentives underlying the network structural and geographical form. In order to address this issue, this paper presents a logit estimation of the determinants of the choices made by firms concerning their R&D partners. In this aim, a precise definition of acquaintance is required. Using the 6th FP database, two firms are considered to be connected if they have submitted a research project to the European Commission together. Our logit estimation highlights two main determinants of the probability to collaborate: firm individual features on the one hand and network effects on the other hand. Among individual features, firms' research potential and their absorptive capability display a significant and positive impact, while their size and leading position in FP projects reduce the probability to collaborate. Regarding network effects, all the indicators of firm position in the network (number of direct and indirect partners, social distance between firms) do affect the probability to collaborate. By contrast, we found no evidence of spatial effects. Then, social distance would matter more than geographical distance.

In spite of these encouraging results, this study is only a first stage in the characterisation of the formation of R&D partnership and in the evaluation of their determinants. Several points need to be improved, concerning both the empirical and the theoretical specifications.

From an empirical point of view, the implications drawn from the study are somewhat restricted due to the scope of the data. In this analysis we have limited ourselves to a specific definition of the network. We focus only on multiprojects industrial partners. Extending the analysis to "one-project" firms or to non-industrial partners (such as research centres or Further Education institutions), we may find a somewhat different picture than we do here. In particular, the incentives to cooperate might differ depending on the type of actors and research centres or universities which might play an intermediary role in the dynamic of cooperation among firms.

Moreover, the sectoral dimension has not been addressed. We have not considered the difference in cooperation behaviours depending on sectors or technological fields. Many previous studies however, have shown significant differences in the R&D and innovation processes among sectors (Powell and al.,1996; Walker and al., 1997). So it would be interesting to broaden our database to other thematic priorities of the European Commission so as to distinguish trans-sectoral regularities from sector-specific cooperation dynamics.

Finally, the geographical dimension is still rather roughly considered here (contiguity at the country level). The data has been processed at the NUTS 2 and 3 levels for further developments on the geographical dimension of cooperation. Finer definitions of the geographical proximity allowing a better measure of the accessibility to partners through space should also be investigated. In particular work is in progress consisting of introducing a measure of distance between the entities according to the centroïdes of regions (NUTS3 level). We will also consider the degree of geographical concentration among the partners.

From a theoretical point of view, future research should address a number of highly relevant questions. Future research should address a number of these issues, using a more appropriate theoretical framework. Two issues are of specific interest:

First, it is clear that we need a new theoretical framework where the incentives to cooperate are defined in a way that better matches the specific case for the proposal of collective R&D projects. Indeed, the assumption set in the cooperative links literature is that agents (or firms) collaborate by pairs. It is however, not really the case in our context, since firms decide to participate to one project with a set of other firms (and some other institutional agents). Furthermore, the literature about the formation of coalitions can not be used in our context, since firms can participate to several projects, and so can belong to different coalitions (it follows that in our case, the collection of the projects is not a partition of the set of firms). It seems that the most useful tools to deal with this kind of network formation belong to the hypergraph theory, where nodes (firms in our context) can belong to different hyper edges (group of firms).

Second, if the structures of networks have huge implications on innovativeness performances and if these structures result from the decision of decentralised interacting actors, we now face another main issue, namely, the characterization of stable (equilibrium) networks compared to the socially efficient ones. Depending on the incentives schemes which stimulate the formation of networks and on the definition of the global efficiency or social value of networks, we can find different ways of dealing with this issue in the literature (Jackson, Wolinski, 1996). What is particularly interesting is that, if there is a conflict between stable networks and efficient networks, then, one can look at the devices that public authorities can use to resolve or mitigate this conflict. Given the importance lately attached by numerous local, national or European authorities to public policies favouring the development of scientific and technological collaborations among R&D actors, it is certainly worth confronting this theoretical issue to actual data.

References:

Barabasi A.L., Albert R. (1999), Emergence of scaling in random networks, *Science*, 286, 509 – 512.

Bala V., Goyal S. (2004), A non-cooperative model of network formation, *Econometrica*, (68), 1181 – 1229.

Bloch F., and Dutta B. (2005), Communication Networks with Endogenous Link Strength, working paper, March.

Borgati S.P. (2005), Centrality and network flow, *Social network*, 25, 55-71.

Breschi and Lissoni (2006), Mobility and social networks: localized knowledge spillovers revisited, *ADRES conference*, Saint-Etienne.

Carayol, N, Roux P., (2005), Self-organizing innovation networks: When do Small Worlds emerge ?, *European Journal of Economic and Social Systems* 18 (2), 307-332.

Cantner and Graf (2006), The network of innovators in Jena: an application of social network analysis, *Research Policy*, 35: 463-480.

Cowan, R. and N. Jonard (2004), Network Structure and the Diffusion of Knowledge" *Journal of*

economic Dynamics and Control, vol 28: 1557-1575

Fafchamps M., van der Leij M.J., Goyal S (2006), Scientific networks and co-authorship, Discussion paper, Février.

Fleming et al. (2004) SmallWorlds and Innovation. Harvard Business School, Mimeo.

Goyal S., van der Leij M., Moraga J.L. (2006), Economics: An Emerging Small, *Journal of Political Economy*, 2, 403-438,

Granovetter,M. (1973), The strength of weak ties, *American Journal of Sociology*, 78, 1360–1380.

Granovetter, M. (1983), The strength of weak ties: a network theory revisited. *Sociological Theory*, 1, 203–233.

Hagedoorn J. (2002) Inter-firm partnerships: an overview of major trends and patterns since 1960, *Research policy*, 31,-492.

Jackson, M., Wolinsky A. (1996), A strategic model of economic and social networks, *Journal of Economic Theory*, 71, 44-74.

Jaffe et al . (1993), Geographic localization of knowledge spillovers as evidenced by patent citations, *Quaterly Journal of Economics*, 10:577-598.

Newman M.E.J. (2004), Coauthorship networks and patterns of scientific collaboration, National Academy of Science USA 101, 5200-5205.

Ponds, Van Oort and Frenken (2006), The geographical and insitutional proximity of scientific collaboration networks, *ADRES Conference*, Saint-Etienne.

Singh J.H. (2005), Collaborative networks as determinants of knowledge diffusion patterns, *Management Science*, 51(5), pp.756-770.

Van der Leij M., Goyal S. (2006), Strong ties in a small world, working paper, May.

Wagner and Leydesdorff (2005), Network structure, self-organisation, and the growth of international collaboration in science, *Research Policy*

Wassserman S., Faust K. (1994) Social network analysis. Methods and applications, *Cambridge University Press*.

Figure 3

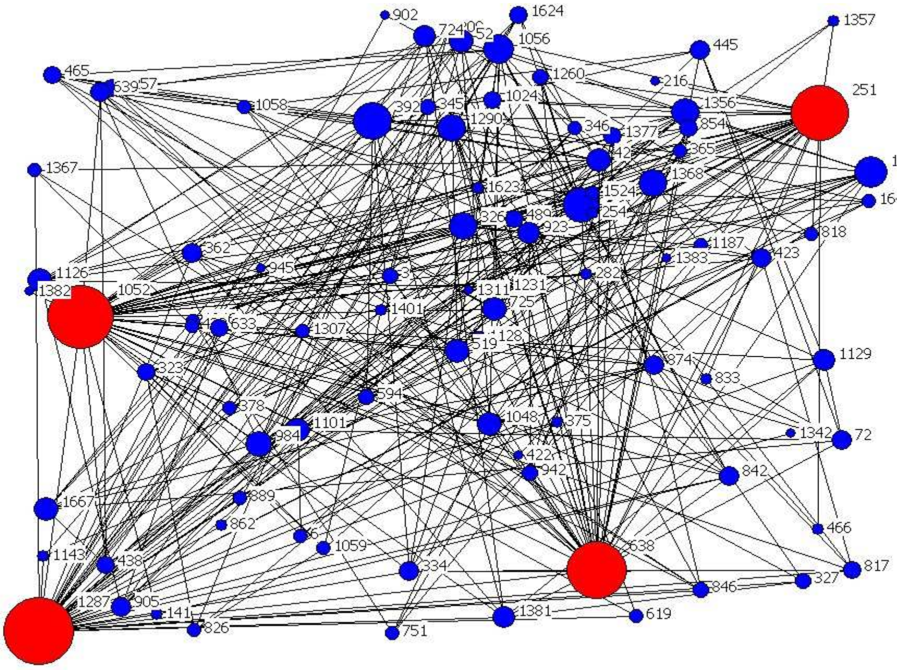


Table 9: Descriptive statistics of the variables

	Mean	Std.Dev.	Minimum	Maximum	NumCases
MProject*	1.417	0.408	1.098	3.332	9591
MBudget*	17.161	0.757	14.492	19.556	9591
Msharebud	7.775	4.035	0	29.934	9591
DProject*	0.804	0.859	0	3.367	9591
DBudget*	16.720	1.424	0	19.590	9591
Dsharebud	6.093	5.381	0	33.109	9591
COMPROJ	0.469^{E-02}	0.859^{E-01}	0	4	9591
Netdist	0.028	0.110	0	1	9591
MPart1	38.23	27.06	4	244.5	9591
Mpart2	903.83	174.48	190.5	1342	9591
Mpart3	1435.18	12.81	1325	1442	9591
RatioPart	84.32	41.5	11.38	519.25	9591

* *variable expressed in logarithm.*