



**HAL**  
open science

# Using Affiliation Networks to Study the Determinants of Multilateral Research Cooperation Some empirical evidence from EU Framework Programs in biotechnology

Cilem Selin Hazir, Corinne Autant-Bernard

► **To cite this version:**

Cilem Selin Hazir, Corinne Autant-Bernard. Using Affiliation Networks to Study the Determinants of Multilateral Research Cooperation Some empirical evidence from EU Framework Programs in biotechnology. 2012. halshs-00697556

**HAL Id: halshs-00697556**

**<https://shs.hal.science/halshs-00697556>**

Preprint submitted on 15 May 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

W P 1212

**Using Affiliation Networks to Study the Determinants of  
Multilateral Research Cooperation : Some empirical evidence  
from EU Framework Programs in biotechnology**

Cilem Selin Hazir, Corinne Autant-Bernard

Mai 2012

**GATE Groupe d'Analyse et de Théorie Économique Lyon-St Étienne**

93, chemin des Mouilles 69130 Ecully – France

Tel. +33 (0)4 72 86 60 60

Fax +33 (0)4 72 86 60 90

6, rue Basse des Rives 42023 Saint-Etienne cedex 02 – France

Tel. +33 (0)4 77 42 19 60

Fax. +33 (0)4 77 42 19 50

Messagerie électronique / Email : [gate@gate.cnrs.fr](mailto:gate@gate.cnrs.fr)

Téléchargement / Download : <http://www.gate.cnrs.fr> – Publications / Working Papers

# Using Affiliation Networks to Study the Determinants of Multilateral Research Cooperation

## Some empirical evidence from EU Framework Programs in biotechnology

Cilem Selin HAZIR<sup>1</sup>, Corinne AUTANT-BERNARD<sup>1</sup>

UMR-CNRS 5824, GATE Lyon Saint-Etienne, University of Saint-Etienne

6 rue Basse des Rives, 42023, Saint-Etienne Cedex 2, France

e-mails: [cilem.selin.hazir@univ-st-etienne.fr](mailto:cilem.selin.hazir@univ-st-etienne.fr), [corinne.autant@univ-st-etienne.fr](mailto:corinne.autant@univ-st-etienne.fr)

### Abstract

*This paper studies multilateral cooperation networks among organizations and work on a two-mode representation to study the decision to participate in a consortium. Our objective is to explain the underlying processes that give rise to multilateral collaboration networks. Particularly, we are interested in how heterogeneity in organizations' attributes plays a part and in the geographical dimension of this formation process. We use the data on project proposals submitted to the 7<sup>th</sup> Framework Program (FP) in the area of Life sciences, Biotechnology and Biochemistry for Sustainable Non-Food. We employ exponential random graph models ( $p^*$  models) (Frank and Strauss, 1986; Wasserman and Pattison, 1996) with node attributes (Agneessens et al., 2004), and we make use of extensions for affiliation networks (Wang et al., 2009). These models do not only enable handling variability in consortium sizes but also relax the assumption on tie/triad independence. We obtained some preliminary results indicating institutional types as a source of heterogeneity affecting participation decisions. Also, these initial results point out that organizations take their potential partners' participations in other projects into account in giving their decision; organizations located in the core European countries tend to participate in the same project; the tendency to preserve the composition of a consortium across projects and the tendency of organizations with the same institutional type to co-participate are not significant.*

**JEL classification:** L14, O31, O38

**Keywords:** Multilateral R&D collaboration, affiliation networks, exponential random graph models, geographical dimension of networks, biotechnology.

---

<sup>1</sup> Université de Lyon, Université Jean Monnet, F - 42023 Saint-Etienne, France.  
CNRS, GATE Lyon-St Etienne, UMR no. 5824, 6 rue basse des rives, 42023 Saint-Etienne, France.  
Phone: +33(0) 477 421 966 Fax: +33(0) 477 421 950.

## 1. Introduction

As corroborated by empirical evidence, research is being carried out more and more in cooperation with others (Hagedoorn, 2002; Wuchty et.al. 2007), creating relational spaces in which knowledge is not only exchanged but also co-created. As one particular type of cooperation one can observe research and development (R&D) project networks, where multiple actors come together in the form of consortiums around an activity plan, budget and a schedule. Indeed, promotion of such collaborations constitutes a particular dimension of innovation policies both at the national and the European level with considerable shares in the direct public support for R&D. To illustrate, in the Seventh Framework Programme (FP7) initiated by the European Commission for 2007-2013, 32.4 billion Euros has been allocated to the Cooperation Program, which is more than half of the budget. Hence, it becomes a matter of interest to know how different combinations of a variety of actors come together to exchange and co-create knowledge; what are the consequences of knowledge flows in such relational spaces in terms of regional disparities in innovation and human capital; and what are the implications for conceptualizing the extent and the meaning of the geographical dimension of externalities.

A certain number of studies have been developed in the field of Economics, which may help to answer these questions. The theoretical literature follows the seminal work of Jackson and Wolinski (1996) regarding the determinants of cooperation and how they impact the structure of networks (Goyal et al., 2006). Johnson and Gilles (2000) and Carayol and Roux (2007) introduced the geographical dimension into this connection model by making the costs of creating connections dependent upon the agents' spatial location links. For reasonably large parameter intervals, and within the network equilibrium, these two models both generate "small world" networks, that is to say, networks characterized by connections with very high rates of proximity connections and a few long-distance connections. Thus, the models predict that geographical proximity should be a central determining factor for collaborative choices. Furthermore, in these two models, the authors show that, in network equilibrium, distant connections may be too few given that which would be socially necessary.

From an empirical perspective, the literature on geography of innovation includes a number of studies addressing the formation of research collaborations. These studies exhibit a high degree of variation in terms of scope, definition of nodes (organization or region) and methodological choice, but they all focus on explaining the determinants of bilateral

cooperation based on one-mode network representations (Autant-Bernard et.al., 2007; Paier and Scherngell, 2008; Scherngell and Barber, 2009; Lata and Scherngell, 2010; Broekel and Hartog, 2011).

However, in multilateral cooperation the nature of the cooperation decisions is different than the one in bilateral cooperation, suggesting a different context to investigate network formation. The reason is that in multilateral cooperation organizations make a single decision on whether to make research together or not rather than making individual decisions for each organization that is likely to be in the consortium. To the best of our knowledge, Frachisse (2010) makes the first attempt to study multilateral cooperation by analyzing the formation of triads in a one-mode representation of a network emerging from FP6 participations. He acknowledges the fact that consortiums may be of any size but due to the difficulty of testing all possible combinations through logit and probit models, he focuses on formation of triads and assumes that formations of triads are independent of each other.

Hence, in this study we focus on multilateral cooperation networks among organizations and work on a two-mode representation to study the decision to participate in a consortium. Our objective is to explain the underlying processes that give rise to multilateral collaboration networks. Particularly, we are interested in how heterogeneity in organizations' attributes plays a part and in the geographical dimension of this formation process. Therefore, different organizations would play different role in the diffusion on knowledge within collaboration networks and their location would affect the ability of knowledge to flow through space. We use the data on project proposals submitted to the 7<sup>th</sup> Framework Program (FP) in the area of Life sciences, Biotechnology and Biochemistry for Sustainable Non-Food. We employ exponential random graph models ( $p^*$  models) (Frank and Strauss, 1986; Wasserman and Pattison, 1996) with node attributes (Agneessens et al., 2004), and we make use of extensions for affiliation networks (Wang et al., 2009). These models do not only enable handling variability in consortium sizes but also relax the assumption on tie/triad independence.

Our preliminary findings show that in the field of Life sciences, Biotechnology and Biochemistry for Sustainable Non-Food, higher education institutions and research institutions have a higher interest in engaging in multilateral research cooperation as compared to other institutional types. In this field, organizations take their potential partners' participations in other projects into account in giving their decisions and organizations located in the core European countries tend to participate in the same project. The tendency to

preserve the composition of a consortium across projects is not found out to be significant. Finally, the tendency of organizations with the same institutional type to co-participate is not found to be significant.

However, we would like to highlight that these are the initial results of an ongoing study and they reflect the behavior of organizations in a very narrow research theme. Hence, they are far from leading us to generalizations. In spite of this, we think that this work might still suggest a value as it attempts to extend the work on multilateral collaboration networks. It builds up a discussion on formation of such networks based on the economic literature on network formation, inter-organizational collaboration, and proximity dimensions; and it presents some initial empirical results. While doing that it also illustrates the use of an alternative modeling approach for multilateral cooperation networks and points out the additional explanatory capacity that one could get by making use of dependencies among ties rather than ignoring them.

In the sequel, we will discuss the determinants of multilateral cooperation based on previous research. Then, in Section 3, we will start presenting the empirical setting by explaining the sample and representation of the network, then we will explore some of the network properties. Afterwards, we will introduce the model and explain variable definitions. In Section 4, we will present the results and finally, in Section 5, we will conclude with an overall assessment.

## **2. Theoretical Framework**

The participation decisions made by each organization could be conceptualized along two dimensions: factors affecting an organization's interest in engaging in multilateral collaboration, and factors affecting their co-participation decisions. We will elaborate each of these dimensions in the following subsections.

### **2.1. Factors affecting the interest in participation**

An organization might be interested in engaging in multilateral research collaborations for a variety of reasons, all of which are not necessarily technology related like improving its business network or brand reputation. However, due to the fact that the context of collaboration is to make R&D and this context is framed by a project plan (activity scheduling, budgeting, division of work, etc) one could assume a stronger role for technology

related aspects and describe this primal interest mainly as interest in accessing information, knowledge, skills, ideas, financial capacity to realize a research that it could have not achieved on its own, etc.

The level of this interest, however, may vary across different institutional types as their primal roles are different. To illustrate, the interest of a public organization in engaging in research consortiums as a user or a regulator is different than a higher education institution who seeks scientific or technological advancements. Furthermore, profit seeking organizations have appropriability concerns as accession has some associated risks about control on the knowledge. Massard and Mehier (2009) argue that there exists a trade-off in maximizing the knowledge acquired and minimizing the loss of appropriability. Cassiman and Veugelers (2002) address this issue for bilateral cooperation in their study based on Community Innovation Survey (CIS) data for Belgium. Their results show that the probability to cooperate is positively related with the level of strategic protection.

**Hypothesis 1:** Organizations with different organizational types differ in their interest in participating in research consortiums.

Another reason why organizations might differ in their interest in participating in research consortiums might be the specific role played by some local features. On the one hand, a high level of industrialization and a well-organized local innovation system in a place might promote systemic learning and interactive innovation (Cooke et al., 1997) and hence foster the absorption capacity of organizations (Cohen and Levinthal, 1990); i.e. “the ability to recognize the value of new information, assimilate it, and apply it to commercial ends”. On the other hand, the development of all regional forms of information services, technological transfer institutions and communication infrastructure may enhance the circulation of information and hence favor the ability of agents to be aware of potential consortiums.

**Hypothesis 2:** Place matters in organizations interest in participating in research consortiums.

In addition to that, each tie has an associated cost and the decision to collaborate is made after considering the costs and benefits associated with the collaboration (Bala and Goyal 2000). However assessment on costs versus benefits cannot be described as an independent evaluation made individually for each tie. At least two reasons could be suggested to support this argument. First, each organization has constraints on the resources it can allocate for



(collaborative) research; and an organizations' portfolio of projects competes for these resources. Second, the benefits that an organization obtains out of a portfolio of projects might be more than the bare sum of individual benefits. As Powell et. al. (1996) argue, learning through collaborations does not take place independently for each tie; indeed, the portfolio of consortiums offer cross-learning opportunities.

**Hypothesis 3:** Organizations tend to increase the number of projects, but each additional project has costs as well as benefits.

## **2.2. Factors affecting co-participation decisions**

The idea at the root of the well-known connection model by Jackson and Wolinski (1996) bases on the fact that establishing a link between agent  $i$  and agent  $j$  allows agent  $i$  to access not only to the information initially held by agent  $j$  but also to other information to which agent  $j$  has access via his other connections (and vice versa for agent  $j$ ). What is more, in the relationship network the further agent  $i$  is away from agent  $j$ , the less the amount of agent  $j$ 's initial information to which he has access. Hence, an organizations' first-hand accession to the knowledge created in other projects might affect its attractiveness as a potential partner.

**Hypothesis 4:** Organizations take partners' accession to other knowledge sources into account in giving the decision to participate in a research consortium.

In addition to that the proximity literature, which provides an important basis for the theoretical framework of studies on bilateral cooperation, bestows some reasoning to explain for the decision to co-participate. On the one hand, social proximity might be a relevant dimension in the sense that a consortium is not merely a medium for knowledge production but also a medium for social interaction. Social ties can play a role to convey information on possible consortiums, their compositions; and make it easier to identify cooperation alternatives and get into contact with them as shown theoretically and empirically in other contexts (Jackson and Rogers, 2007; Fafchamps et. al. 2010; Autant-Bernard et al. 2007). This may result in co-participations with a subset of old partners in new projects.

Furthermore, social ties may affect co-participation decisions through development of trust. Zand (1972) argues that high trust enables the exchange of ideas more openly; and Zaheer et.al. (1998) suggest that the higher the level of inter-organizational trust the less the cost of

negotiations and conflicts. Also, Uzzi (1996) argues that embedded ties transmit more private and tacit knowledge as compared to the information exchanged at arm's-length. Hence, these may give reasons to preserve relationships with old partners by engaging in new projects with them.

In addition to social proximity, the change in the cognitive distance between the members of a consortium may create reasons for future collaborations. The reason is that the members of a consortium not only increases the level of their common knowledge but also develop skills on communicating and doing research collectively. The literature on cognitive distance and related variety argues that sharing common knowledge is a pre-requisite for understanding each other and benefit from collaboration (Frenken et al., 2007; Nooteboom et al., 2007). In the same line the concept of “*relative absorptive capacity*” proposed by Lane and Lubatkin (1998) points out that an organisation's ability to learn changes from partner to partner. This change results from the degree of similarity in knowledge bases, compensation practices and organisational structures, and organisational problem set. However, too much proximity may also hamper collaboration, by increasing the risk of lock-in and by reducing the opportunities of new knowledge creation. As noticed by Cowan, Jonard and Zimmerman (2007) heterogeneity among agents regarding their knowledge levels decreases as knowledge diffuses within the network, reducing the propensity to generate innovation.

**Hypothesis 5:** Organizations tend to preserve their partners across participations (stability of consortiums) but there is a trade-off between this tendency and the risk of lock-in.

On the other hand geographical proximity can be argued to be playing a role, independently or together with social proximity. This role could be conceptualized along several axes. First, physical proximity could be associated with acquaintance or awareness about organizations nearby. In this regard, agglomerations might be associated with a higher possibility for agents to be identified. Geographical proximity between agents may favor flows of information and knowledge because, as stated by A. Moles (1992) what is close to me (here and now) is more important than what is far (in space and time)<sup>2</sup>. Economic agents will thus pay more attention to what happens in the geographical area in which they are located. Firms are indeed federated by local institutions which causes joint meeting opportunities. It may be official

---

<sup>2</sup> « Ce qui est proche pour moi (ici et maintenant) est plus important que ce qui est lointain (ailleurs, autrefois, plus tard) » A. Moles, « Vers une psycho-géographie », in A. Bailly, R. Ferras and D. Pumain, ed. (1992).

bodies (Chambers of Commerce and Industry, departmental or regional union branches, etc.). But also more informal structures such as local associations (with objectives in terms of training, environmental protection, for example) and whose only reason for grouping result from their belonging to the same geographical area. Thus, companies may meet because of geographic proximity, while at a distance they would not have the opportunity. The same applies to relations between firms and public research organizations. Proximity to universities allows for many interactions, likely to ease collaboration between agents (use of academic consultants, recruitment of young doctors, library use, informal communications at conferences or social activities, etc.). Second, geographical proximity may act as a facilitator for face-to-face interactions and promote transmission of tacit knowledge (Feldman and Florida, 1994; Anselin, Varga, and Acs, 2000; Balland, 2009); and hence, increase the amount and the scope of knowledge that can be accessed. Geographical proximity may also facilitate cross-fertilization of ideas (Feldman and Florida 1994), pointing out a higher potential of knowledge that could be co-created. In addition to that, it may enable timely inflows of information (Feldman, 1993) or reduce the cost of collaboration (Hoekman et. al., 2009).

However, the specific role played by distance is difficult to distinguish from the one played by some local features. While being spatially proximate might mean sharing similar local features, leading to spatial dependence in collaboration decisions; the local features may also affect the interest in collaboration in the sense that the local innovation system may play a crucial part on the ability of agents to be aware of potential collaborators as well as of their attributes. Some places benefit from well organized information systems promoting both systemic learning and interactive innovation (Cooke et al., 1997). More generally, the development of all regional forms of information services, technological transfer institutions and communication infrastructure may foster the circulation of information and hence favor the ability to co-participate in R&D networks.

**Hypothesis 6:** There is a tendency among organizations located in core regions to participate in the same project (see Section 3.2.3 for definitions).

Another aspect of similarity refers to the institutional proximity, which is relevant both at the country or regional level and at the organizational level (Knoeben and Oerlemans, 2006). It refers to the closeness in standards, routines, values, goals, languages, etc., which are “*enabling mechanisms that provide stable conditions for interactive learning*” (Boschma,

2009). Balland (2009) models R&D collaborations in FP6 in the navigation satellite industry as bilateral collaborations and finds out that organizations prefer to interact when they are institutionally proximate. On the other hand, Ponds et. al. (2007) claim that geographical proximity can help overcoming problems resulting from differences among goals in research, institutional backgrounds and constraints.

**Hypothesis 7:** There is a tendency among organizations with the same institutional type to participate in the same project.

Finally, co-participations should have a limit in the sense that the larger the size of the consortium, the harder to coordinate and achieve cross-learning. Hence, trade-offs arise as the unique piece of knowledge or the unique capacity to create knowledge offered by each additional participant have associated costs in terms of project management.

**Hypothesis 8:** Consortium sizes tend to increase, but each additional partner brings in coordination costs as well as learning benefits.

### **3. Empirical Setting**

#### **3.1. Data Description and Construction of the Sample**

The data is provided by the French Ministry of Higher Education and Research on the basis of the European Commission records on the project proposals submitted to the 7<sup>th</sup> Framework Program (FP7). In this raw data set, there is a high degree of variation in the way the name of an organization is recorded for different project proposals it is involved in. The use of data in its raw form, hence, results in either treating the same organization as several different organizations depending on the number of different names under which it is recorded, or treating geographically/institutionally different entities as the same organization as they are registered in the database only with the name of the legal entity that they are a member of. To reduce this ambiguity in organizations, the raw data is processed by EUROLIO (European Localized Innovation Observatory) to match the records that correspond to the same entity by making use of the information on the location of the organization.

The overall dataset includes proposals submitted within the scope of four programs of FP7: cooperation, ideas, people, and capacities. Due to the scope of this study we used the information in cooperation program, by which transnational research is funded by the

Commission in ten different thematic fields. Among those ten fields, we restricted our study to Food, Agriculture and Fisheries, Biotechnology first, and then further to the subtheme called Life sciences, biotechnology and biochemistry for sustainable non-food, which consists of a wide range of subtopics. Among the projects in this field we used small or medium-scale focused research projects to build our sample.

The rationale behind such a restriction is two-folds. First, the thematic fields and their subthemes are not defined from a purely technological point of view, indeed some of them serve for fulfillment of particular socio-economic objectives. Hence, some subthemes are associated with diverse technological areas. Due to the fact that the dynamics of network formation may differ in different knowledge basis, we tried to focus on a subtheme with a narrower technological span. Second, due to our model choice and associated difficulties in estimation which will be explained in Section 3.3, we tried to arrive at a sample with a manageable size. As a result we end up with a sample of 237 projects including 1316 unique participants, which has been proposed to the Commission in response to five different calls issued yearly from 2007 to 2011.

### **3.2. Network Representation and Network Properties**

Using the sample explained in the previous section, we build a two mode network, which consists of two types of nodes and ties among them. In this network, the first set ( $A = \{1,2,3, \dots, n\}$ ) of nodes refers to organizations, and the second set ( $P = \{1,2,3, \dots, m\}$ ) refers to projects. Hence in our network, the set ( $\Omega$ ) of all possible ties connecting each organization in  $A$  to each project in  $P$  is of size  $n \times m$ . We denote a possible tie between an organization  $i \in A$  and a project  $j \in P$ , with the random variable  $Y_{ij}$ , which takes a value of 1 if the tie is realized (meaning that organization  $i$  participated in project  $j$ ), and 0 otherwise. Then we express the overall network as a random vector ( $\mathbf{Y}$ ), which is a collection of tie variables, i.e.  $\mathbf{Y} = \{Y_{ij}\}$ . We denote a realization of this vector with  $\mathbf{y} = \{y_{ij}\}$ .

Table 1 summarizes the size of the network and realization of ties. In the following subsections, we will abridge major properties of organizations and consortiums.

**Table 1.** Network Size

<b>Number of Organizations (<math>n</math>)</b>	1 316
<b>Number of Projects (<math>m</math>)</b>	237
<b>Number of Possible Ties (<math> \Omega  = n \times m</math>)</b>	311 892
<b>Number of Realized Ties</b>	2 153
<b>Average Degree of Organizations (Min and Max)</b>	1.64 (1 and 17)
<b>Average Degree of Projects (Min and Max)</b>	9.08 (1 and 18)
<b>Average Density</b>	0.0069

### 3.2.1. Heterogeneity among Organizations: Difference across Institutional Types

When we consider the distribution of organizations with respect to their institutional types (Table 2); we observe that higher education institutions (HES) appear to be the largest group followed by private commercial organizations (PRC) and research organizations (REC). Public institutions (PUB) and other types of organizations (OTH) together account for only 7.6% of the actors in the network.

However, when we consider the share of each group in the total number of participations, we observe that these shares are not directly proportional to group sizes, signaling a difference in their interest in participating in research consortiums as formulated in Hypothesis 1. While PRC's constitute one third of the organizations in the network, one fourth of the participations are made by them. On the other hand, HESs constitute 39% of the organizations in the network, but almost 46% participations are made by them. The behavior of PUBs and OTHs are similar to PRCs, whereas the behavior of RECs is similar to HESs as depicted in Table 2.

**Table 2.** Breakdown of Organizations and Participations with respect to Institutional Types (%)

	<b>HES</b>	<b>PRC</b>	<b>REC</b>	<b>PUB</b>	<b>OTH</b>	<b>TOTAL</b>
Breakdown of Organizations	39.1	32.6	20.7	1.9	5.7	100.0
Breakdown of Participations	46.3	25.4	22.5	1.6	4.2	100.0

Examining the degree distribution of organizations, this difference is further elicited. While the maximum degree is 17 for HES and REC, it is only 3 for PUB. Furthermore, the distribution of organizations with respect to their degrees differs across institutional types. While the majority in all institutional types tends to have a degree of one, resulting in positively skewed degree distributions; the degree of variation in degrees and skewness of the degree distribution differs across different institutional types as indicated in Table 3.

**Table 3.** Differences in Degree Distributions across Institutional Types

	<b>HES</b>	<b>PRC</b>	<b>REC</b>	<b>PUB</b>	<b>OTH</b>	<b>Overall Network</b>
Minimum degree	1	1	1	1	1	1
Maximum degree	17	5	17	3	6	17
Mean degree	1.9	1.3	1.8	1.4	1.2	1.64
Std. Dev. of the Actor Degree Dist.	1.82	0.69	1.74	0.69	0.68	1.48
Skewness of the Actor Degree Dist.	1.41	0.98	0.98	0.31	0.29	4.65

Finally, Table 3 indicates that although there are organizations that engage in as many consortiums as 17, on the average the project portfolio of an organization is of size 2. This observation goes hand in hand with Hypothesis 3 that bases on the trade-off between multi-connectivity versus cost of connections.

### **3.2.2. Heterogeneity among Organizations: Difference across Locations**

Our data set includes information on the location of organizations; however while this information is available up to NUTS3 level for EU countries, regional information is not available for non-member countries. Organizations from non-member countries constitute around 19% of all organizations in the network. For countries with a small land area, assuming that the country consists of a single region might be reasonable but for countries with very large land areas this assumption results in underestimation of the geographical coverage at the regional level. With this limitation Table 4 summarizes the geographical coverage the network under study.

The table includes a classification of regions as core, periphery, and other. We call regions located in Austria, Belgium, France, Germany, Italia, the Netherlands, Switzerland, United Kingdom as the core regions. We call regions in EU-member countries other than core regions as the periphery regions and classify the rest as other. In our network, 49% of the organizations are located in core countries and 53% of all participations are made by this group. On the other hand, 32% of organizations are located in the periphery and the participations made by them also constitute 32% of all participations. Finally, we observe that the situation for other regions is the reverse of the core. These observations are in the same line with Hypothesis 2.

**Table 4.** Geographical Coverage of the Network

<b>Number of NUTS3 Regions Involved *</b>	490
<b>Number of Countries Involved</b>	81
<b>Share of Organizations Located at Core Regions (%)</b>	48.9
<b>Share of Organizations Located at Periphery Regions (%)</b>	32.0
<b>Share of Organizations Located at Other Regions (%)</b>	19.1
<b>Share of Participations from Core Regions (%)</b>	53.0
<b>Share of Participations from Periphery Regions (%)</b>	31.9
<b>Share of Participations from Other Regions (%)</b>	15.1

\* underestimated as region information is not available for non-member countries.

### 3.2.3. Composition of Consortiums

In this section we will try to describe the composition of consortiums to explore the co-participation decisions. To start with, in Table 5 we present the share of consortiums including co-located participants. Indeed, the geographical coverage of consortiums is not free of the design of FP program, since the Commission sets the minimum conditions on the consortium size and location of participants. According to these conditions:

*“at least three legal entities must participate, each of which must be established in a Member State or associated country, and no two of which may be established in the same Member State or associated country”*

*(Regulation(EC) No 1906/2006; Article 5/(1))*

Breakdown 1 in Table 5 reveals that, after satisfying the condition for three members stated above, the consortiums tend to increase co-locations rather than increasing the country variety. Breakdown 2 in the same table reveals further that these co-locations tend to be as narrow in geographical scope as NUTS3 level. Breakdown 3 however goes hand in hand with Hypothesis 6 as in 95% of the consortiums there are at least one pair of organizations located in the core.

Finally, when we consider the composition of consortiums in terms of institutional types, we observe that consortiums that are homogeneous in terms of institutional types are rare (less than 2%). As indicated in Table 6, approximately 60% of all consortiums bring together organizations from three different institutional types. Indeed, majority of these consortiums include higher education institutions (HES), private commercial organizations (PRC) and research organizations (REC) as the three different institutional types.



**Table 5.** Consortiums with Co-located Partners

	<b>Count</b>	<b>Share (%)</b>
Total number of consortiums	237	100
Breakdown 1:		
Including at least two participants co-located at the country level	227	95.8
Including participants each from a different country	10	4.2
Breakdown 2:		
Including at least two participants co-located at the NUTS3 level*	142	59.9
Including participants each from a different NUTS3 region**	95	40.1
Breakdown 3:		
Including at least two participants located in a core region	225	94.9
Including participants at most one located in a core region	12	5.1

\* overestimated, \*\* underestimated

**Table 6.** Composition of Consortiums with respect to Institutional Types of Participants

	<b>Count</b>	<b>Share (%)</b>
with single institutional type	4	1.7
with 2 institutional types	27	11.4
with 3 institutional types	148	62.4
with 4 institutional types	53	22.4
with 5 institutional types	5	2.1
<b>Total</b>	<b>237</b>	<b>100</b>

\* 118 of which are consortiums among HES, PRC, and REC.

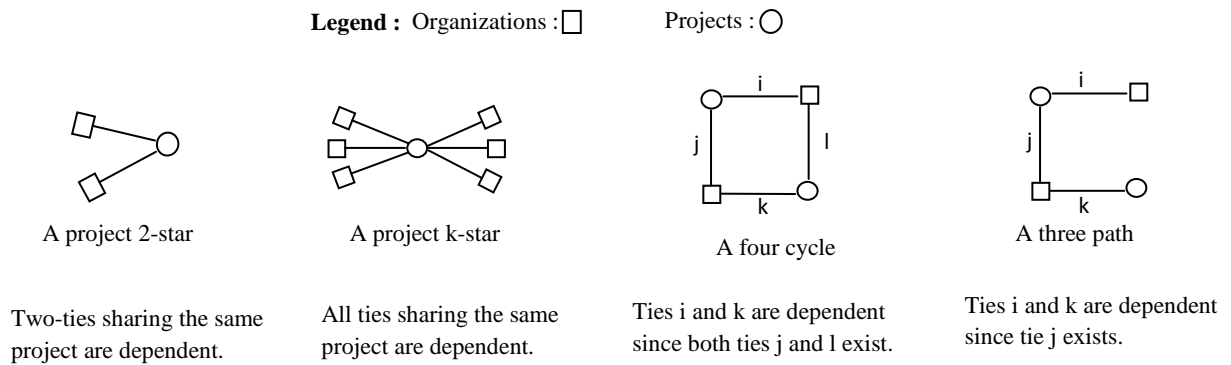
### 3.3. The Model

One of the key conclusions that could be derived from the previous discussion on the determinants of multilateral consortium is that the decision made by an organization to participate in a consortium is not an independent decision. Indeed, even the simple fact that a consortium is realized as the joint result of participation decisions made by its members, corroborates this. As a result, models based on tie independence like Bernoulli random graphs (Erdős and Renyi, 1959) or  $p_1$  models (Holland and Leinhardt, 1972) suggest a limited capacity to study such networks.

Exponential random graph models ( $p^*$  models) (Frank and Strauss, 1986; Wasserman and Pattison, 1996), however, not only avoids an assumption on tie independence but also permit taking a wide range of tie dependencies into account. Indeed, dependency assumptions lie at the heart of  $p^*$  models as they express the probability of observing a particular network configuration in terms of the local configurations (neighborhoods) that emerge as a result of these dependence assumptions. For the statistical and mathematical foundations of  $p^*$  models we refer readers to the joint probability of a Markov field or to the extensions of statistical

mechanics of Gibbs to the study of networks by Park and Newman (2004), and to the Hammersely Clifford Theorem (Besag, 1974) proving the Gibbs-Markov equivalence.

One class of dependencies that could be handled through a  $p^*$  model is called the Markov dependence (Frank and Strauss, 1986). According to Markov dependence assumption, two ties not sharing a node are conditionally independent. Another class, called realization dependence assumption (Pattison and Robbins, 2002) allows a high degree of freedom in incorporating tie dependencies. Realization dependence assumptions refer to the assumptions where two ties are assumed to be dependent conditional on the realization of other ties. Wang et al. (2009) extended these definitions to two-mode networks and Agneesens et.al (2004) incorporated node attributes to two-mode networks. Figure 1.a and Figure 1.b provides examples for the two classes of dependence assumptions for two-mode networks and the resulting local configurations (neighborhoods).



**Figure 1.a.** Examples to Markov Dependence

**Figure 1.b.** Examples to Realization Dependence

ERGM has the following general form (Robins et.al, 2007):

$$P(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa}\right) \exp\left\{\sum_Q \eta_Q g_Q(\mathbf{y})\right\}$$

Where the following definitions hold:

- $P(\mathbf{Y} = \mathbf{y})$  is the probability of observing a particular network  $\mathbf{y}$ .
- $\kappa$  is a normalizing constant assuring the probabilities given by this distribution adds up to 1:

$$\kappa = \sum_{\mathbf{y} \in \mathcal{Y}} \left(\exp \sum_Q \eta_Q g_Q(\mathbf{y})\right)$$

- $\eta_Q$  is the parameter corresponding to the local configuration (neighborhood)  $Q$ .
- $g_Q(\mathbf{y})$  is the network statistic corresponding to the local configuration (neighborhood)  $Q$ . In a homogeneous model, for a given type of neighborhood  $Q$ , which is a collection of isomorphic neighborhoods  $q$ ,  $g_Q(\mathbf{y})$  is given by:

$$g_Q(\mathbf{y}) = \sum_{q \in Q} \left( \prod_{Y_{ij} \in q} y_{ij} \right)$$

Estimating an ERGM means finding the parameter values that maximizes the probability of observing the network of interest. However, the term  $\kappa$  in the model includes a summation over the set of all possible graphs with the same size, which is impossible to compute even for small networks as the cardinality of this set is quite high. There are two main techniques suggested to cope with this difficulty: Pseudo-Likelihood Estimation (Straus and Ikeda, 1990) and Markov Chain Monte Carlo Maximum Likelihood Estimation (MCMCMLE) (Snijders, 2002). Geyer and Thompson (1992) state that in cases when there is strong dependence among the ties, PLE may overestimate the dependence parameters. Similarly, Robins et. al. (2007) point out to that limitation and argue that the use of this method stems from its practical convenience but indeed it “does not have a principled basis”. They suggest the use of Monte Carlo method whenever it is possible. Finally, Wang et. al. (2009) provide empirical evidence on the performance of the two estimation techniques for bipartite graphs and propose that MCMCMLE should be the preferred method for bipartite graphs.

### 3.4. Definition of Variables

We defined a total of 12 variables to test the hypothesis presented in Section 2. We list and explain them below and in Table 7 we provide a classification relating variables to hypothesis and local configurations (neighborhoods).

- **L**: Total number of edges in the network. This is indeed a default variable in ERGM as it is the smallest neighborhood that makes up a network. It may be interpreted as the average interest by an organization in participating in research consortiums.
- **hes\_rA**: Total number of edges formed by higher education institutions.
- **prc\_rA**: Total number of edges formed by private enterprises.
- **rec\_rA**: Total number of edges formed by research institutions.

- $corehes\_rA$ : Total number of edges formed by the higher education institutions located in core regions.
- $coreprc\_rA$ : Total number of edges formed by private enterprises located in core regions.
- $corerec\_rA$ : Total number of edges formed by the research institutions located in core regions.
- $Ksa$ : Alternating organization k-stars is defined as in Wang et. al. (2009) by making use of the number of organization k-stars ( $g_{astar_k}(\mathbf{y})$ ):

$$g_{Ksa}(\lambda_s, \mathbf{y}) = \sum_{k=2}^n (-1)^k \frac{g_{astar_k}(\mathbf{y})}{\lambda_s^{k-2}} \quad \text{where} \quad g_{astar_k}(\mathbf{y}) = \sum_{i=1}^n \binom{y_{i+}}{k}$$

and  $\lambda_s$  is the weighting parameter,  $|A| = n$ ,  $i \in A$ ,  $y_{i+}$  denote the degree of organization  $i$ . In our estimations we used the default value of  $\lambda_s$ , which is equal to 2.

- $core\_tsoA2$ : Total number of 2 paths centered at projects among organizations located in core regions. We define a binary variable  $c_i$  which takes a value of 1 if the organization  $i \in A$  is located in the core, 0 otherwise. Then:

$$g_{2pcore}(\mathbf{y}) = \sum_{i=1}^{r-1} \sum_{r=2}^n \sum_{j=1}^m y_{ij} y_{rj} c_i c_r$$

where  $i, r \in A$ ,  $j \in P$ ,  $|A| = n$ ,  $|P| = m$ .

- $L3$ : Three paths are defined as in Wang et. al. (2009) by means of the number of two-paths ( $L_{2ir}(\mathbf{y})$ ) between a particular organization pair ( $i$  and  $r$ ):

$$g_{L3}(\mathbf{y}) = \sum_{i=1}^{r-1} \sum_{r=2}^n L_{2ir} (y_{i+} + y_{r+} - 2) \quad \text{where} \quad L_{2ir}(\mathbf{y}) = \sum_{j=1}^m y_{ij} y_{rj}$$

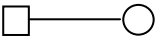
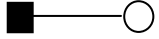
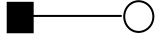
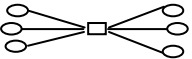
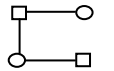

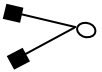
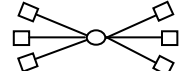
and  $i, r \in A$ ,  $j \in P$ ,  $|A| = n$ ,  $|P| = m$ ,  $y_{i+}$  and  $y_{r+}$  denote the degree of organization  $i$  and  $r$ , respectively.

**Table 7. Summary of Hypothesis and Variable Definitions**

Organization with or without attribute: □

Organization with binary attribute: ■

Projects: ○

<b>Hypothesis related with factors affecting the interest in participation</b>			
<b>Statement</b>	<b>Variable Name</b>	<b>Definition</b>	<b>Shape of the Local Configuration</b>
Default hypothesis: Organizations are interested in participating in research consortiums.	L	Total number of edges	
H1: Organizations with different organizational types differ in their interest in participating in research consortiums.	hes_rA	Total number of edges by higher education institutions	
	prc_rA	Total number of edges by private institutions	
	rec_rA	Total number of edges by research institutions	
H2: Place matters in organizations interest in participating in research consortiums.	corehes_rA	Total number of edges formed by the higher education institutions located in core regions.	
	coreprc_rA	Total number of edges formed by the private institutions located in core regions.	
	corerec_rA	Total number of edges formed by the research institutions located in core regions.	
H3: Organizations tend to increase the number of projects, but each additional project has costs as well as benefits.	Ksa	Alternating actor (organization) k-stars	
<b>Hypothesis related with the determinants of co- participation decisions</b>			
H4: Organizations take partners' accession to other knowledge sources into account in giving the decision to participate in a research consortium.	L3	Total number of three paths	
H5: Organizations tend to preserve their partners across participations (stability of consortiums) but there is a trade-off between this tendency and the risk of lock-in.	Kca	Actor centered alternating k-two paths	
H6: There is a tendency among organizations located in core regions to participate in the same project (see Section 3.2.3 for definitions).	core_tsoA2	Total number of 2 paths centered at projects among organizations located in core regions.	
H7: There is a tendency among organizations with the same institutional type to participate in the same project.	instype_match_2pA	Total number of 2 paths centered at projects with matching institutional types.	
H8: Consortiums tend to have large sizes but each additional member brings additional costs as well as benefits.	Ksp	Alternating project k-stars	

- Kca: Organization centered k-two paths is defined as in Wang et. al. (2009) similar to Ksa:

$$g_{Kca}(\lambda_t, \mathbf{y}) = \lambda_t \sum_{i=1}^{r-1} \sum_{r=2}^n \left\{ 1 - \left(1 - \frac{1}{\lambda_t}\right)^{L_{2ir}} \right\}$$

where  $\lambda_t$  is the weighting parameter,  $r \in A$ ,  $|A| = n$ , and  $L_{2ir}(\mathbf{y})$  is calculated as in L3. In our estimations we used the default value of  $\lambda_t$ , which is equal to 2.

- instype\_match\_2pA: Total number of 2 paths centered at projects with matching institutional types. We define a categorical variable  $t_i$  which indicates the institutional type of the organization  $i \in A$ . Then:

$$g_{2pinstype}(\mathbf{y}) = \sum_{i=1}^{r-1} \sum_{r=2}^n \sum_{j=1}^m y_{ij} y_{rj} I\{t_i = t_r\}$$

where  $i, r \in A$ ,  $j \in P$ ,  $|A| = n$ ,  $|P| = m$ , and  $I$  is an indicator function which returns a value of 1 if the logical statement is true, and 0 otherwise.

- Ksp: Alternating project k-stars is defined as in Wang et. al. (2009) by making use of the number of project k-stars  $g_{pstar_k}(\mathbf{y})$ :

$$g_{Ksp}(\lambda_s, \mathbf{y}) = \sum_{k=2}^m (-1)^k \frac{g_{pstar_k}(\mathbf{y})}{\lambda_s^{k-2}} \quad \text{where} \quad g_{pstar_k}(\mathbf{y}) = \sum_{j=1}^m \binom{y_{j+}}{k}$$

And  $\lambda_s$  is the weighting parameter,  $|P| = n$ ,  $j \in P$ ,  $y_{j+}$  denote the degree of project  $j$ . In our estimations we used the default value of  $\lambda_s$ , which is equal to 2.

#### 4. Estimation Results and Discussion

There are two types of measures to assess the quality of the parameter estimates given by the MCMCMLE procedure. The first one is related to the quality of convergence, and the latter refers to the goodness of fit (GOF). The quality of convergence is measured by t-ratios calculated to check whether the estimate of the parameter vector ( $\widehat{\eta}_Q$ ) is capable of producing a graph distribution centered at the observed network (Wang et. al, 2009). To do that a large number of simulation iterations are carried out, out of which the expected value and the

standard deviation of the graph statistics are computed (Snijders, 2002). Then, the t-ratios are calculated by the following formula:

$$t_Q = \frac{E[g_Q(\mathbf{Y})] - g_Q(\mathbf{y})}{\hat{\sigma}(g_Q(\mathbf{Y})|\hat{\eta}_Q)}$$

Snijders (2002) suggests that the convergence of an ERGM model is excellent if  $|t_Q| \leq 0.1$ ; good if  $0.1 < |t_Q| \leq 0.2$ ; and poor if  $0.2 < |t_Q| \leq 0.3$  for  $\forall Q$ .

After arriving at a converged model, the next step is to conduct a simulation analysis to check how far the center of the graph distribution generated by the model is to the observed network. To do that a number of graph statistics (which do not only include the ones included in the model but also some other structural network statistics that might be of interest) are collected from the simulated graph distribution and their t-ratios are computed. For graph statistics that are included in the model all t-ratios should be less than or equal to 0.1. For the other statistics, Wang et. al. (2009) accepts t-ratios smaller than 2.0 as a heuristic approach. They also suggest the use of Mahalanobis distance ( $d_M$ ) as a measure for the distance from the center of the graph distribution generated from the model to the observed network. Hence, a smaller  $d_M$  indicates a better goodness of fit.  $d_M$  is defined as follows:

$$d_M = \sqrt{(G(\mathbf{y}) - \mu)^T \Sigma^{-1} (G(\mathbf{y}) - \mu)}$$

where  $G(\mathbf{y}) = [g_1(\mathbf{y}), g_2(\mathbf{y}), \dots, g_Q(\mathbf{y})]$  is the vector of observed network statistics;  $\mu = [\mu_1, \mu_2, \dots, \mu_i]$  is the vector of corresponding means obtained by the simulated graphs and  $\Sigma$  is the covariance matrix.

We conduct all our estimations and goodness-of-fit analysis using “BPNet”, which is an extension of the PNet programme (Wang et. al. 2006) and approximates the maximum likelihood estimate through a Monte Carlo procedure based on Robbins-Monro algorithm. The results we obtained so far are abridged in Tables 8 to 10.

Table 8 displays the parameter estimates for the model specifications that have excellently converged; i.e.  $|t_Q| \leq 0.1$  for  $\forall Q$ . One of the striking results is revealed by the comparison of Model 1 and Model 2. Model 1 describes the simplest specification where ties are assumed to be independent and the average interest by an actor in engaging in a consortium is investigated. In this model, the negative and statistically significant parameter estimate for  $L$

means that networks with higher number of edges are less likely to occur. Model 2, however, tests the entire hypothesis related to the factors affecting the interest in participation. It reveals that the interest of higher education institutions (HES) and research organizations (REC), in engaging in consortiums is higher than that of public institutions (PUB), appear to be the largest group followed by private commercial organizations (PRC), and other types of organizations Others (OTH); i.e. networks with higher number of edges by HESs and RECs are more probable. Furthermore, this interest is higher for those HESs and RECs located in the core. Although both models have converged excellently, comparison of GOF results presented in Table 9 and Table 10 reveal that Model 1 dominates Model 2. This indeed points out a particularity and at the same time a difficulty of ERGM in the sense that adding new types of neighborhoods to the specification does not necessarily improve GOF.

**Table 8.** Parameter Estimates for Converged Specifications (standard deviations in parenthesis)

Effects	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
<b>L</b>	-4.970* (0.022)	-4.466* (0.133)	-5.412* (0.093)	-5.481* (0.095)	-5.418* (0.097)	-5.506* (0.091)	-5.472* (0.102)
<b>hes_rA</b>		0.494* (0.133)	0.240* (0.093)	0.389* (0.094)	0.302* (0.123)	0.497* (0.118)	0.405* (0.134)
<b>prc_rA</b>		-0.027 (0.148)	-0.028 (0.107)	-0.001 (0.099)	0.006 (0.117)	0.056 (0.098)	0.091 (0.127)
<b>rec_rA</b>		0.450* (0.147)	0.215* (0.096)	0.318* (0.099)	0.247* (0.116)	0.367* (0.097)	0.325* (0.116)
<b>corehes_rA</b>		0.583* (0.084)	0.278* (0.059)		0.275* (0.059)		0.171** (0.102)
<b>coreprc_rA</b>		0.107 (0.116)	0.056 (0.086)		0.059 (0.084)		-0.051 (0.117)
<b>corerec_rA</b>		0.371* (0.120)	0.185* (0.084)		0.183* (0.081)		0.074 (0.114)
<b>Ksa</b>		-0.897* (0.074)					
<b>L3</b>			0.005* (0.002)	0.002 (0.002)	0.004* (0.001)	0.004** (0.002)	0.005** (0.002)
<b>Kca</b>			-0.005 (0.039)	0.043 (0.094)		0.013 (0.047)	-0.005 (0.053)
<b>core_tsoA2</b>				0.036* (0.007)		0.036* (0.007)	0.021 (0.016)
<b>instype_match_2pA</b>					-0.015 (0.018)	-0.026 (0.021)	-0.025 (0.022)

\* significant at 95% level, \*\* significant at 90% level.

Another point to mention about Table 8 is that we had convergence problems for specifications including Ksp and also with Ksa. We observed in GOF analysis that specifications without these terms (Model 3 to Model 7) are well capable of reproducing Ksp statistics in the observed network but poorly capable of generating as many Ksa as in the



network under study (Table 9). This means that our converged specifications do not allow us to explain the number of projects in which each organization is involved.

Finally, examining Model 3 to Model 7 in Table 8 reveals that each of these models tests and validates a different subset of hypothesis. While Table 9 shows that these models perform similarly and better than Model 1 and Model 2 in generating a graph distribution that is closer to the observed one; Table 10 displays that Model 6 has the smallest Mahalanobis distance.

Model 6 tests the hypothesis 1, 4, 5, 6, and 7; and among those it validates hypothesis 1, 4, and 6. The negative and statistically significant parameter estimate for edges imply that network with less number of edges are more probable. The positive and statistically significant parameter estimates for *hes\_rA* and *rec\_rA* mean that the interest by HES and REC is different than that of PUB and OTH; and networks with higher number of edges by HES and REC are more probable. The parameter estimate for *L3* shows that organizations take their potential partners' participations in other projects into account. Hence, networks with a higher number of three-paths are more probable. Nevertheless, the magnitude of the parameter is very small. Finally, the positive and statistically significant parameter estimate for *core\_tsoA2* implies that organizations located in the core tend to participate in the same project; therefore networks with higher number of co-participations by organizations located in the core are more probable.

Examining goodness of fit (GOF) results for Model 6, we see that it is well capable of reproducing a number of graph statistics that are not included in the model. Well reproduced graph statistics include the number of organization 2 stars, project 2 stars, project 3-stars, alternating project k-stars, project centered k-2 paths; edges by HES, PRC and REC located in the core; standard deviation of the degree distribution of organizations, and skewness of degree distribution of projects. However, it poorly reproduces mainly two properties of the actual network. Indeed these properties are to some extent interrelated and poorly reproduced by the other models as well.

One of the two not-well reproduced properties is clustering and the poor performance of the model in this respect is reflected by the high t-ratios of four-cycles and global clustering. The other property is the very long and very thin tail of the actor degree distribution with a mass having a degree of 1. High t-ratios for organization 3-stars, alternating organization k-stars,

and skewness of degree distribution of organizations points out to the failure in reproducing this property.

**Table 9.** GOF Results: t-ratios for major graph Statistics\* (statistics included in the model in bold)

Effects	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
<b>L</b>	<b>0.013</b>	<b>-0.021</b>	<b>0.032</b>	<b>0.066</b>	<b>0.050</b>	<b>-0.025</b>	<b>0.016</b>
Organization 2-stars	4.649	12.333	0.735	0.897	0.684	0.711	0.727
Project 2-stars	-0.551	-0.685	-0.666	-0.719	-0.560	-0.704	-0.647
Organization 3-stars	35.162	63.421	12.331	14.180	12.081	12.801	12.512
Project 3-stars	-0.966	-1.118	-1.189	-1.303	-1.029	-1.221	-1.151
<b>L3</b>	2.827	8.345	<b>-0.001</b>	<b>0.076</b>	<b>0.039</b>	<b>-0.030</b>	<b>0.004</b>
Four-cycles	43.028	58.537	23.451	26.634	24.803	23.731	24.261
<b>Ksa</b>	-3.823	<b>-0.011</b>	-3.730	-4.090	-3.832	-3.862	-3.796
<b>Ksp</b>	-0.024	-0.066	-0.010	0.021	0.012	-0.065	-0.023
<b>Kca</b>	3.315	10.465	<b>0.013</b>	<b>0.082</b>	-0.050	<b>-0.033</b>	<b>-0.008</b>
Project centered k-2 paths	-0.992	-1.263	-0.946	-1.025	-0.852	-0.985	-0.929
hes_rA	5.444	<b>0.003</b>	<b>0.037</b>	<b>0.092</b>	<b>0.051</b>	<b>-0.042</b>	<b>-0.005</b>
prc_rA	-5.958	<b>-0.019</b>	<b>0.070</b>	<b>0.035</b>	<b>0.083</b>	<b>0.006</b>	<b>0.017</b>
rec_rA	1.712	<b>-0.014</b>	<b>-0.011</b>	<b>0.009</b>	<b>-0.034</b>	<b>-0.026</b>	<b>0.024</b>
corehes_rA	8.036	<b>-0.039</b>	<b>0.005</b>	1.138	<b>0.019</b>	1.027	<b>-0.022</b>
coreprc_rA	-4.370	<b>0.001</b>	<b>0.078</b>	-0.797	<b>0.052</b>	-0.815	<b>0.041</b>
corerec_rA	3.001	<b>0.012</b>	<b>0.012</b>	0.033	<b>-0.031</b>	-0.028	<b>0.019</b>
core_tsoA2	3.689	0.864	0.311	<b>0.053</b>	0.395	<b>-0.023</b>	<b>0.018</b>
Instype_match_2pa	1.873	0.133	-0.216	-0.215	<b>0.033</b>	<b>-0.045</b>	<b>0.016</b>
Std dev. of degree dist. of organizations	7.684	18.954	1.260	1.486	1.144	1.280	1.275
Skewness of degree dist. of organizations	48.351	43.088	29.070	31.270	27.637	30.154	28.521
Std dev. of degree dist. of projects	-2.400	-2.286	-3.514	-3.527	-3.125	-3.432	-3.428
Skewness of degree dist. of projects	0.040	0.038	-0.500	-0.652	-0.310	-0.611	-0.491
Global clustering	42.659	42.001	35.292	38.812	37.026	36.127	36.427

\* calculated on the basis of 1000 graphs sampled using an interval of 2 million iterations and 100.000 iterations for burn-in.

**Table 10.** Comparison of Mahalanobis Distances

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
$d_M^*$	486.1	1139.1	192.4	198.1	209.9	185.9	194.7

\* calculated on the basis of 1000 graphs sampled using an interval of 2 million iterations and 100.000 iterations for burn-in.

On the other hand, the observed number of four-cycles in the network is only 431 and the global clustering is very low (0.046). Similarly, the share of organizations with a degree higher than or equal to 3 is 14%. Furthermore an analysis of the components of four cycles

and organizations with high degrees reveal that an important part of four cycles are created by organizations with high degrees, meaning that the statistics that are not well reproduced do not correspond to a very frequent and widely encountered behavior in the network. Yet, this does not eradicate the need for further experimenting to mature this work.

Until now, we have implemented several strategies to improve the fit in these two aspects. These include introducing subthemes as project attributes, introducing time as an attribute both for projects and organizations (ex: using information on FP calls to introduce project start times and organizations' entry time to the network as attributes), introducing the number of four-cycles (with or without attributes), replacing Kca with tendency to extend co-participations across multiple projects (project centered k-two paths), etc. However, no further improvements could be achieved via these strategies.

## **5. Conclusions**

In this study, we consider that multilateral research cooperation defines a particular context to study network formation as organizations get connected to a number of organizations through a single decision on joining to a consortium. While such networks are among the main tools of innovation policies, investigating the mechanisms of their formation is a key issue, providing us with information on the way knowledge diffuses as well as on the factors hampering this diffusion. Based on the economic literature on network formation, inter-organizational collaboration, and proximity dimensions; we presented some reasoning to explain for the formation of these networks. As an empirical application, based on the data on proposals submitted to FP7 on a specific sub-theme we analyzed a multilateral cooperation network among organizations using a two-mode representation and exponential random graph models.

One of our objectives was to investigate the heterogeneity among agents in their interest in participating in such networks. Although the results we obtain are specific to a particular research, they still suggest a general conclusion in the sense that they point out to two types of heterogeneity sources: institutional types and location. We consider that the study of the effects of these two sources in different research areas would yield important results for the design of relevant public policies.

Another main objective of our study was to shed some light upon the geographical dimension of the formation of these networks. Nevertheless, our work suffered from two basic limitations: unavailability of regional information for non EU countries, and the EC regulation on minimum conditions on the consortium size and location of participants. While the former hindered a more precise exploratory analysis at NUTS3 level; the latter led to the cancellation of a hypothesis on the effect of co-location on co-participation. Still, we managed to draw several empirical results through a core-periphery perspective. The descriptive analysis revealed that in 95% of the consortiums there is at least one organization pair located in the European core countries. The econometric part, also, revealed some initial results on the tendency of organizations located in the core to co-participate.

Another conclusion that could be drawn from this study is that models which can handle tie dependence rather than ignoring it suggest an additional explanatory capacity. We illustrated this by using ERGM and testing hypothesis on stability of consortium compositions across projects and the effect of partners' accession to other knowledge sources on co-participations. Hence, we would like to highlight the high degree of freedom that is suggested by the realization dependence assumption (Pattison and Robbins, 2002) in studying complex economic phenomenon.

On the other hand, we would also like to point out the limitations in estimating this kind of models. While, the span of theoretical issues that could be addressed by these models is quite broad, coming up with a proper specification and arriving at a converging specification is not trivial. One of the main difficulties occurs during the Monte Carlo process in specifying parameter values to produce realizations that cover the observed values of the network statistics (Handcock, 2003). Another major difficulty is that inclusion of more complex local structures does not assure convergence (Wang et. al. 2012). Furthermore, as stated in Wang et. al. (2012), several specifications may converge but they display different capacity to reproduce the properties of the observed network; unlike generalized linear models addition of parameters does not necessarily improve goodness of fit.

Finally, as we have mentioned several times before throughout the text, the work we presented here corresponds to the initial part of an ongoing study. Like many other colleagues, in this part we considered the network under study as a single realization in time as if all actors make all their decisions simultaneously. Obviously, an organization makes some of its decisions simultaneously, and some at different time instances. To the best of our

knowledge, Lata and Scherngell (2010) and Balland (2009) addressed this issue for bilateral cooperation networks with a particular interest on the geographical dimension. In addition to that Hanneke et. al., (2010) and Krivitsky et. al. (2010) proposed temporal extensions to ERGM to study the evolution as a discrete time Markov process. In this respect, we think that integrating a temporal aspect to study the evolution of a multilateral cooperation network is promising.

## References:

- Agneessens et. al. (2004) Choices of Theatre Events: p\* Model for Affiliation Networks With attributes, *Metodološki zvezki*, 1(2): 419-439.
- Anselin L., Varga A. and Acs Z. (2000) Geographical Spillovers and University Research: A Spatial Econometric Perspective, *Growth and Change*, 31: 501-515.
- Autant-Bernard, C. et al., (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(3): 495–519.
- Bala V and Goyal S (2000) A non-cooperative model of network formation, *Econometrica* 68: 1181-1229.
- Balland P.A. (2009), “Proximity and the Evolution of Collaboration Networks Evidences from R&D Projects within the GNSS industry,” presented at 6th Journées de la Proximité, Poitiers, France.
- Besag J. (1974) Spatial Interaction and the Statistical Analysis of Lattice Systems, *Journal of the Royal Statistical Society. Series B (Methodological)*, 36 (2): 192-236.
- Boschma R.A. (2005), Proximity and Innovation: A critical Assessment, *Regional Studies*, 39(1), 61-74.
- Broekel T. and Hartog M. (2011) Explaining the Structure of Inter-organizational Networks Using Exponential Random Graph Models: Does Proximity Matter?, *Papers in Evolutionary Geography* No.11.07.
- Carayol N. and Roux P. (2007) The Strategic Formation of Inter-Individual Collaboration Networks: Emprical Evidence from Co-invention Patterns, *Annales d'Economie et de Statistiques*, 87-88: 275-301.
- Cassiman B. and Veugelers R. (2002) R&D Cooperation and Spillovers: Some Emprical Evidence from Belgium, *The American Economic Review*, 92(4): 1169-1184.
- Cohen W.M. and Levinthal D.A. (1990) Absorptive Capacity: A New Perspective on Learning and Innovation, *Administrative Quarterly*, 35(1), Special Issue: Technology, Organizations, and Innovation: 128-152.
- Cooke P., Uranga M.G., Etxebarria G. (1997) Regional innovation systems: institutional and organisational dimensions. *Research Policy* 26: 475-491
- Cowan R., Jonard N. and Zimmermann J-B. (2007) Bilateral Collaboration and the Emergence of Innovation Networks, *Management Science*, 53(7): 1051-1067,
- Erdős, P. and Renyi, A. (1959), On random graphs I, *Publicationes Mathematicae Debrecen* (6), pp. 290-297.
- Fafchamps M., Goyal S. and van der Leij M.J. (2010) Matching and Network Effects, *Journal of the European Economic Association*, 8(1): 203-231
- Feldman M.P. (1993) An Examination of the Geography of Innovation. *Industrial and Corporate Change* 2, no. 3.
- Feldman M.P and Florida R. (1994) The Geographic Sources of Innovation: Technological Infrastructure and Product Innovation in the United States, *Annals of the Association of American Geographers*, 84(2): 210-229.

- Frachisse D. (2010) L'analyse des réseaux sociaux appliquée aux PCRDT: Enjeux méthodologiques et outils pour l'évaluation, PhD Thesis. CREUSET, Université de Saint-Etienne, France.
- Frank O. and Strauss D. (1986) Markov Graphs, *Journal of American Statistical Association*, 81(395): 832-842.
- Frenken K., Van Oort F. and Verburg T. (2007) Related Variety, Unrelated Variety and Regional Economic Growth, *Regional Studies*, 41(5): 685-697.
- Geyer C. J. and Thompson E. A. (1992) Constrained Monte Carlo Maximum Likelihood for Dependent Data, *Journal of the Royal Statistical Society, Series B (Methodological)*, 54 (3): 657-699.
- Goyal S., van der Leij M.J., Moraga J.L. (2006) Economics : An emerging small world?, *Journal of Political Economy*, 2:403-438.
- Hagedoorn J. (2002), Inter-firm R&D Partnerships: an Overview of Major Trends and Patterns Since 1960, *Research Policy*, 31, 477-492.
- Handcock M.S. (2003) Assessing Degeneracy in Statistical Models of Social Networks, Working Paper no.39, Centre for Statistics and Social Sciences, University of Washington.
- Hanneke S.; Fu W. and Xing E.P.(2010) Discrete Temporal Models of Social Networks, *Electronic Journal of Statistics*, 4: 585-605.
- Hoekman J., Frenken K. and van Oort F. (2009) The Geography of Collaborative Knowledge Production in Europe, *The Annals of Regional Science*, 43(3): 721-738.
- Holland P.W. and Leinhardt S. (1981) An Exponential Family of Probability Distributions for Directed Graphs, *Journal of American Statistical Association*, 6(373): 33-50.
- Jackson O.M. and Rogers B.W. (2007) Meeting Strangers and Friends of Friends: How Random are Social Networks?, *The American Economic Review*, 97(3): 890-915.
- Jackson M. O. and Wolinsky A. (1996) A strategic Model of Social and Economic Networks, *Journal of Economic Theory*, 71: 44-74.
- Johnson C. and Gilles R. (2000) Spatial social networks, *Review of Economic Design*, 5: 273-299.
- Knoben J. and Oerlemans L.A.G. (2006), "Proximity and Inter-organizational Collaboration: A Literature Review," *International Journal of Management Reviews* 8, no. 2: 71-89.
- Krivitsky P.N. and Handcock M.S. (2010) A Separable Model for Dynamic Networks, arXiv:1011.1937v1.
- Lata R. and Scherngell T. (2010) The Spatio-Temporal Distribution of European R&D Networks: Evidence Using Eigenvector Spatially Filtered Spatial Interaction Models, Available at SSRN: <http://ssrn.com/abstract=1720945>
- Lane P.J. and Lubatkin M. (1998) Relative Absorptive Capacity and Inter-organizational Learning, *Strategic Management Journal*, 19: 461-477.
- Massard N. and Mehier C. (2009) Proximity and Innovation Through an "Accessibility of Knowledge" Lens, *Regional Studies*, 43: 77-88.
- Nooteboom B., et. al. (2007) Optimal Cognitive Distance and Absorptive Capacity, *Research Policy*, 36(7): 1016-1034.
- Paier M. and Scherngell T. (2008) Determinants of Collaboration in European R&D Networks: Empirical Evidence from a Binary Choice Model Perspective, NEMO Working Paper 10.

- Park J. and Newman M.E.J. (2004) The Statistical Mechanics of Networks, *Physical Review E*, 70(6).
- Pattison P. and Robins G. (2002) Neighborhood-based models for social networks, *Sociological Methodology*, 32: 301-337.
- Ponds R., van Oort F. and Frenken K. (2007) The Geographical and Institutional Proximity of Research Collaboration, *Papers in Regional Science*, 86(3): 423-443.
- Powell W.W., Koput K.W. and Smith-Doerr L. (1996) Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology, *Administrative Science Quarterly*, 41(1): 116- 145.
- Regulation (EC) No 1906/2006 of the European Parliament and of the Council of 18 December 2006 laying down the rules for the participation of undertakings, research centres and universities in actions under the Seventh Framework Programme and for the dissemination of research results (2007-2013)
- Robins G. et. al. (2007) An Introduction to Exponential Random Graph ( $p^*$ ) Models for Social Networks, *Social Networks*, 29: 173–191.
- Scherngell T. and Barber M.J. (2009) Spatial Interaction Modelling of Cross-Region R&D Collaborations: Empirical Evidence from the 5th EU Framework Program, *Papers in Regional Science*, 88(3): 531–546.
- Snijders T.A.B. (2002) Markov Chain Monte Carlo Estimation of Exponential Random Graph Models, *Journal of Social Structure*, 3(2).
- Straus D. and Ikeda M. (1990) Pseudolikelihood Estimation for Social Networks, *Journal of American Statistical Association*, 85(409): 204-212.
- Uzzi B. (1996) The Sources and Consequences of Embeddedness for the Economic Performance of Organizations: The Network Effect, *American Sociological Review*, 61(4): 674-698.
- Wang P. et. al. (2006) PNet: A Program for the Simulation and Estimation of Exponential Random Graph Models. School of Behaviour Science, University of Melbourne.  
URL <http://www.sna.unimelb.edu.au/pnet/pnet.html>.
- Wang P. et.al. (2009) Exponential Random Graph ( $p^*$ ) Models for Affiliation Networks, *Social Networks*, 31/ 12-25.
- Wang P., Pattison, P, and Robins G. (2012) Exponential Random Graph Model Specifications for bipartite networks- A dependence Hierarchy, *Social Networks*, doi:10.1016/j.socnet.2011.12.004.
- Wasserman S. and Pattison P. (1996) Logit Models and Logistic Regression for social Networks: I. An Introduction to Markov Graphs and  $p^*$ , *Psychometrika*, 61(3): 401-425.
- Wuchty S., Jones B.F. and Uzzi B. (2007) The Increasing Dominance of Teams in Production of Knowledge, *Science*, 316: 1036 – 1039.
- Zaheer A., McEvily B. and Perrone V. (1998) Does Trust Matter? Exploring the effects of Interorganizational and Interpersonal Trust on Performance, *Organizational Science*, 9(2): 141-159.
- Zand D.E. (1972) Trust and Managerial Problem Solving, *Administrative Science Quarterly*, 17(2): 229-239.