

Barber, Michael J.; Fischer, Manfred M.; Scherngell, Thomas

Conference Paper

The Community Structure of R&D Cooperation in Europe. Evidence from a social networks perspective

50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Barber, Michael J.; Fischer, Manfred M.; Scherngell, Thomas (2010) : The Community Structure of R&D Cooperation in Europe. Evidence from a social networks perspective, 50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/118939>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The Community Structure of R&D Cooperation in Europe.

Evidence from a social network perspective

Michael J. Barber¹, Manfred M. Fischer² and Thomas Scherngell¹

¹Foresight and Policy Development Department,
Austrian Institute of Technology (AIT) Vienna, Austria

²Institute for Economic Geography and GIScience,
Vienna University of Economics and Business (WU Wien) Vienna, Austria

Abstract. The focus of this paper is on pre-competitive R&D cooperation across Europe, as captured by R&D joint ventures funded by the European Commission in the time period 1998-2002, within the 5th Framework Program. The cooperations in this Framework Program give rise to a collaborative network, with network nodes representing actors (i.e. organizations including firms, universities, research organizations and public agencies) and network edges representing R&D projects. With this construction, participating actors are linked only through joint projects. We formally describe and analyze the network from a social network perspective that shifts attention to the detection and analysis of the community structure within the network. Distinct communities within networks may be loosely defined as groups of actors such that there is a higher density of relations within groups than between them. In this study, we attempt to detect communities of actors solely on the basis of the relational structure within the network, and to characterize and differentiate the identified network communities by means of information-theoretic methods, community-specific profiles and the location of their major actors. We expect the results to enrich our picture of the European Research Area (ERA) by providing new insights into the global and local structures of R&D cooperation across Europe.

JEL Classification: C02, C49, L14, O39, O52

Keywords: R&D cooperation, European Framework Program, large-scale networks, decomposition of networks, network community identification, social network analysis

*Paper to be presented at the 50th European Congress of the Regional Science Association,
19-23 August 2010, Jönköping, Sweden*

1 Introduction

Knowledge production takes place within a complex web of interactions among firms, universities and research institutions (see, for instance, Fischer 2001). Long viewed as a temporary, inherently unstable organisational arrangement, R&D networks have become the norm rather than the exception in modern innovation processes (Powell and Grodal 2005). In the recent past, regional, national and supranational STI policies have emphasized supporting and fostering linkages between innovating actors (for a discussion of major international examples, see Caloghirou et al. 2002). At the European level, the main STI policy instruments are the European Framework Programmes (FPs) which promote an integrated European Research Area (ERA). The FPs support pre-competitive R&D projects, creating a pan-European network of actors performing joint R&D.

In this paper, we examine pre-competitive European¹ R&D cooperations from a social network perspective. A social network perspective focuses not on the individual social actors, but on the broader interaction contexts within which the actors are embedded. Social network analysis explicitly assumes that actors participate in social systems connecting them to other actors, whose relations comprise important influences on one another's behaviours. Central to network analysis are identifying, measuring, and testing hypotheses about the structural forms and substantive contents of relations among actors. This distinctive structural-relational emphasis sets social network analysis apart from individualistic, variable-centric traditions in the social sciences (Knoke and Young 2008).

The importance of social network analysis rests on two underlying assumptions. First, structural relations are often more important for understanding observed behaviours than are attributes of the actors. Second, social networks affect actors' perceptions, beliefs and actions through a variety of structural mechanisms that are socially constructed by relations among them. Direct contacts and more intensive interactions dispose actors to better information, greater awareness, and higher susceptibility to influencing or being influenced by others. Indirect relations through intermediaries also bring exposure to new ideas and access to useful

¹ R&D networks constituted under the heading of the FPs have recently attracted a number of empirical studies. Scherngell and Barber (2009 and 2010) focus on the geography of pre-competitive R&D networks across European regions by using data on joint research projects of FP5. Breschi and Lissoni (2004) employ a social network perspective to analyse R&D collaborations with the objective to unveil the texture of the European Research Area (ERA).

resources that may be acquired through interactions with others. Networks provide complex pathways for assisting or hindering flows of information and knowledge.

In this work we explore pre-competitive R&D cooperation across Europe from a social network perspective. We focus on a network derived from R&D joint ventures funded by the European Commission in the time period 1998-2002, within the 5th Framework Program (FP5). FP5 gives rise to a network consisting of a set of nodes or vertices representing 25,839 formal organizations, such as firms, universities and research organizations, connected together by links or edges, representing cooperation in FP5 projects. The objective is to detect and describe the community structure of this network, using the recently introduced label-propagation algorithm (LPA). Communities are loosely defined as partitioning the nodes or vertices into groups such that there is a higher density of links within them than between them. The definition is based on comparing intra-group density to inter-group sparseness. The popularity of density-based grouping is due to the likelihood that actors within communities share common properties and/or play similar roles within the network. This is the motivation for analysing network communities in general (see Fortunato 2010 for a recent review) and for analysing European R&D network communities in particular.

The paper is organised as follows. *Section 2* describes the community identification problem based on the concept of modularity. *Section 3* introduces the LPA approach to identify communities in the network under consideration. The LPA was originally presented operationally, with communities defined as the outcome of a specific procedure. In this work, we consider an equivalent mathematical formulation, in which community solutions are understood in terms of optima of an objective function. *Section 4* differentiates the identified communities by developing community-specific profiles using social network analysis and geographic visualisation techniques. *Section 5* concludes with a summary of the main results, some policy implications and a short outlook.

2 The community-identification problem

A network of R&D cooperation can be viewed in several ways. One of the most useful views is as a graph consisting of vertices (nodes) and edges (links). Let V be a set of vertices, representing actors participating in FP5, and E be a set of vertex pairs or edges from $V \times V$, representing participation in a joint FP5 project. The two sets together are a graph $G=(V, E)$.

In a simple graph, all pairs $\{u, v\} \in E$ are distinct and $\{u, u\} \notin E$. Given a partition $V=V_1+V_2$ where no edges exist between pairs of elements within V_1 or V_2 , then G is said to be bipartite.

We shall consider simple graphs on a large finite set $V=\{1, 2, \dots, n\}$. The number of edges in the graph is denoted by m , the number of edges incident on a vertex $i=1, \dots, n$ is called the degree k_i . The topology of the graph is encoded in the $n \times n$ adjacency matrix \mathbf{A} with elements

$$A_{ij} = \begin{cases} 1 & \text{if } \{i, j\} \in E \\ 0 & \text{otherwise.} \end{cases} \quad i, j=1, \dots, n \quad (1)$$

In many real world networks there are large inhomogeneities in the degrees, reflecting a high level of order and structure. The degree distribution is highly skewed; many vertices with low degrees coexist with some vertices with high degrees. The distribution of edges may be both globally and locally inhomogeneous, with high concentration of edges within specific groups of nodes, and low concentration between these groups. This feature of real world networks is called community structure.

There are different ways to define the community-identification problem. The most prominent formulation is based on the concept of modularity, a measure that evaluates the quality of a partition of a graph into subsets of vertices in comparison to a null model. Formally, the modularity Q is defined as

$$Q \equiv \frac{1}{2m} \sum_{i,j} (A_{ij} - P_{ij}) \delta(g_i, g_j) \quad (2)$$

with the Kronecker delta term

$$\delta(g_i, g_j) = \begin{cases} 1 & \text{if } g_i = g_j \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where g_i and g_j denote the community groups to which vertices i and j are assigned, respectively, and P_{ij} denotes the probability in the null model that an edge exists between vertices i and j . Thus, the modularity Q is – up to a normalization constant – defined as the number of edges within communities minus those expected in the null model.

The standard choice of the null model is that proposed by Newman and Girvan (2004) and consists of a randomized version of the actual graph, where edges are rewired at random, under the constraint that each vertex i keeps its degree k_i .

Denoting $E(A_{ij})$ by P_{ij} and assuming further that

$$P_{ij} = p_i p_j \quad (4)$$

then

$$P_{ij} \equiv \frac{k_i k_j}{2m}. \quad (5)$$

With this choice for P_{ij} , the modularity becomes

$$Q \equiv \frac{1}{2m} \sum_{i,j} \left(A_{ik} - \frac{k_i k_j}{2m} \right) \delta(g_i, g_j). \quad (6)$$

The goal now is to find a division of the vertices into communities such that the modularity Q is high. An exhaustive search for a decomposition is out of question. Even for moderately large networks there are far too many ways to decompose them into communities.

In this study, we account for the bipartite character of the network in question. Bipartite networks have additional constraints which can be reflected in the null model. For bipartite graphs, the null model should be modified to reproduce the characteristic form of bipartite adjacency matrices (see Barber and Clark 2009 for more details)

$$A = \begin{bmatrix} O_{n_1 \times n_1} & \tilde{A}_{n_1 \times n_2} \\ (\tilde{A}^T)_{n_2 \times n_1} & O_{n_2 \times n_2} \end{bmatrix} \quad (7)$$

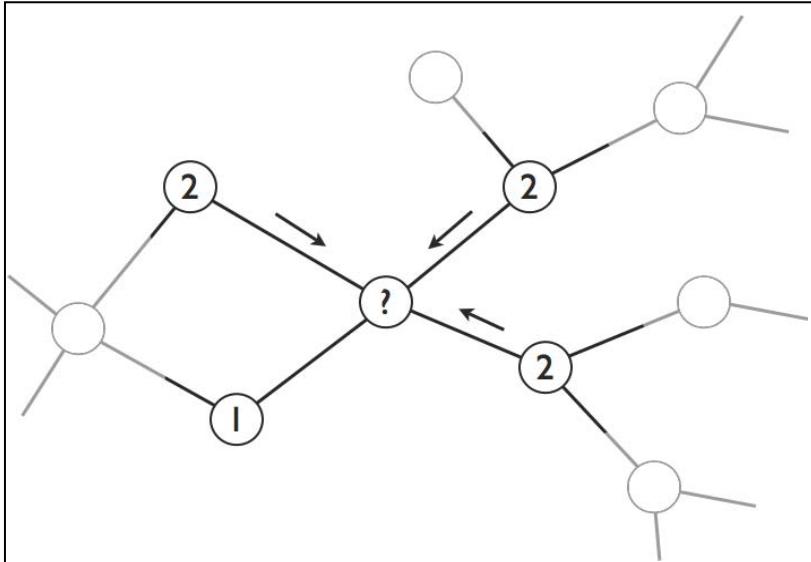
with n_1 and n_2 denoting the number of vertices in V_1 and V_2 , respectively, and $n = n_1 + n_2$.

$O_{i \times j}$ is the all-zero matrix with i rows and j columns.

3 A label-propagation algorithm for maximizing (bipartite) modularity

We use a label propagation algorithm (LPA) introduced by Raghavan et al. (2007) to identify community groups. In this approach, community assignments are described by labels assigned to the network vertices. Beginning with a unique label for each vertex, labels are dynamically updated until a stable assignment of labels is obtained. Network communities are then taken to be sets of vertices bearing the same labels, with labels propagating from vertices to their neighbours (see Figure 1).

Figure 1: Updating community assignment by propagating labels



We formalize the LPA following the presentation of Barber and Clark (2009), describing the LPA as an optimization problem. We introduce an objective function H , which is just the number of edges linking vertices with the same label, i.e. in the same community group g . This can be expressed formally in terms of the adjacency matrix, giving

$$g'_v = \arg \max_g \sum_{u=1}^n A_{uv} \delta(g_u, g) \quad (8)$$

Label assignment corresponds to selecting a community group g for vertex v that maximizes H , i.e., a label that occurs most frequently among the neighbors of v . Formally, this is

$$H = \frac{1}{2} \sum_{v=1}^n \sum_{u=1}^n A_{uv} \delta(g_u, g_v). \quad (9)$$

It may be that multiple choices of g would produce a maximal H . In such a case, a specific label is selected by keeping the current label if it would satisfy (8) and otherwise taking a label at random that satisfies (9). This excludes non-terminating cycles where a vertex varies between different labels satisfying (9).

To put the label-update rule (9) into effect, we must also define an update schedule. A practical approach suggested by Raghavan et al. (2007) is to update the vertex labels asynchronously and in random order. Multiple updating passes are made through the vertices, continuing until all vertices have labels satisfying (9). This update schedule ensures termination of the search by eliminating cycles where two neighboring vertices continually exchange labels.

The LPA offers a number of desirable qualities. As described above, it is conceptually simple, being readily understood and quickly implemented. The algorithm is efficient in practice. Each relabeling iteration through the vertices has a computational complexity linear in the number of edges in the graph. The total number of iterations is not a priori clear, but relatively few iterations are needed to assign the final label to most of the vertices (over 95% of vertices in 5 iterations, see Raghavan et al. 2007, Leung et al. 2008).

A significant drawback of the LPA is that the objective function H corresponds poorly to our conceptual understanding of communities. In fact, the global maximum in H is trivially obtained by assigning the same label to all vertices, providing no information at all on community structure. Interesting community solutions thus must be located at local maxima in H , but H offers no mechanism for comparing the quality of the solutions. An auxiliary measure, such as the modularity Q , can be introduced to assess community quality. Using modularity, communities found using LPA are seen to be of high quality (Raghavan et al. 2007); label propagation is both fast and effective.

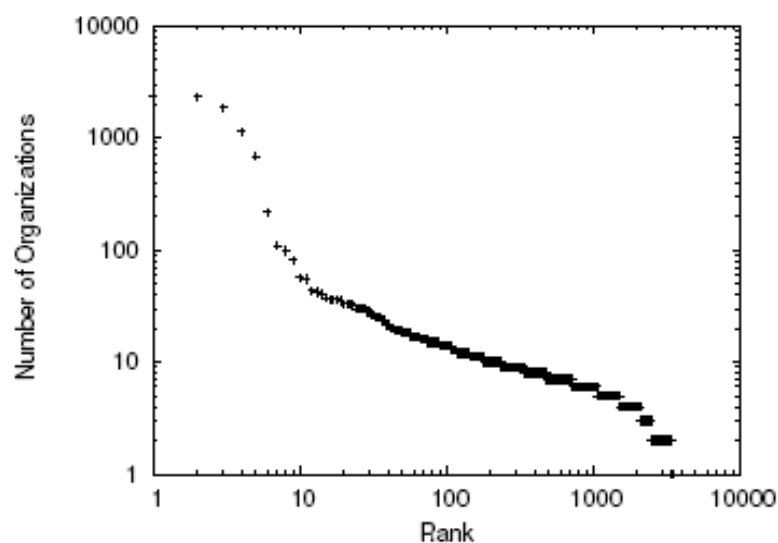
Barber and Clark (2009) have elucidated the connection between label propagation and modularity, showing that modularity can be maximized by propagating labels subject to additional constraints and proposing several variations of the LPA. In this work, we make use

of a hybrid, two-stage label propagation scheme, consisting of LPA_r followed by LPA_b (see Barber and Clark 2009 for details). The first stage, LPA_r, is similar to the original LPA presented above, but with additional randomness to allow the algorithm to avoid premature termination. In the second stage, LPA_b, constraints are imposed on the label propagation so that the algorithm identifies a local maximum in Q_B , a version of the modularity specialized to bipartite networks (Barber 2007).

4 Network communities and topical differentiation

In this section, we use the LPA approach to identify and differentiate communities for the European R&D cooperation network. We develop community-specific profiles to thematically characterize the network communities, and consider their spatial distribution. We identified 3,482 network communities. The communities vary greatly in size, as measured either by the number of organizations in the community or by the number of projects in the community (ranked by size in Figure 2). Most (2,878) communities consist of just a single project with some or all of the participating organizations. In contrast, 20 or more projects are observed in just nine communities, but they contain over a third of the organizations and over half of the projects present in FP5. For the rest of this paper, we will consider only these largest communities (see Table 1 and Figure 3).

Figure 2: Rankings of communities by number of organizations



Thematic differentiation and characterization of the network communities

The communities are identified using only the network structure, but that structure arises from the processes by which organizations organize projects. To gain a better understanding of the nature of the communities, we thus examine the properties of the constituent organizations and projects. We focus particularly on three characteristics: (i) the standardized subject indices (sometimes also referred to as keywords) assigned to the projects by the EU, (ii) the names of the projects, and (iii) the identity of the organizations. By considering these, we find a strong thematic character for the communities. We summarize the community themes concisely in Table 1 and provide additional details below.

As a first step, we gain a basic understanding of the communities by examining their thematic orientation using standardized subject indices assigned to the projects in the community. There are 49 subject indices in total, ranging from *Aerospace Technology* to *Waste Management*; a complete list of subject indices is given by CORDIS (2008). Absolute counts of projects with a particular subject index are uninformative, as the subject indices occur with different frequencies in FP5 projects. More meaningful is to compare the number of projects N_s in a community featuring a subject index S to the number $E[N_s]$ we would expect if the projects were chosen at random from FP5; differences in the values can be tested for statistical significance using a binomial test. In Table 1, we show the most strongly over-represented subject indices for each community, giving the values as a ratio $R_s = N_s / E[N_s]$ of actual occurrences to expected occurrences of the index. The subject indices are strongly suggestive of thematic differentiation between the communities, with communities apparently oriented towards the life sciences, transportation, electronics, and other topics.

Further insight into the communities is gained by examining the project titles, allowing a more specific characterization of their thematic character. Particularly for the larger communities, the titles suggest possible community substructures of more specialized nature; we note the presence of such subnetworks, but do not pursue them further in this work. Using the standardized subject indices and the project titles, we assigned the names as shown in Table 1 to each community.

Table 1: Characterization of communities by thematic orientation

	$R_s > 5$	$5 \geq R_s > 3$	$3 \geq R_s > 1$
Life Sciences	-	Biotechnology***; Life Sciences***; Medicine, Health***; Regional Development***	Agriculture***; Food***; Policies***; Safety***; Scientific Research***; Social Aspects***; Waste Management***
Aquatic Resources	Agriculture***; Resources of the Sea, Fisheries***	Life Sciences***	Economic Aspects***; Environmental Protection***
Electronics	-	Electronics, Microelectronics***; Evaluation*; Telecommunications***	Education, Training***; Forecasting***; Information Processing, Information Systems***; Information, Agriculture*; Environmental Protection***; Measurement Methods*; Regional Development*; Scientific Research*;
Environment	Earth Sciences***; Meteorology***; Standards***	Forecasting***; Resources of the Sea, Fisheries***	Agriculture*; Environmental Protection***; Measurement Methods*; Regional Development*; Scientific Research*;
Sea Transport	Transport***	Safety***	Environmental Protection***
Ground Transport	Energy Storage, Energy Transport***	Fossil Fuels**	Energy Saving***; Environmental Protection*; Materials Technology*; Reference Materials*; Safety***
Aerospace	Aerospace Technology***	Energy Saving***; Energy Storage, Energy Transport***; Renewable Sources of Energy***; Transport***	Industrial Manufacture***; Information Processing, Information Systems***; Other Energy Topics**
Information Processing	Electronics, Microelectronics***; Legislation, Regulations***; Mathematics, Statistics***; Policies***	-	Information Processing, Information Systems***

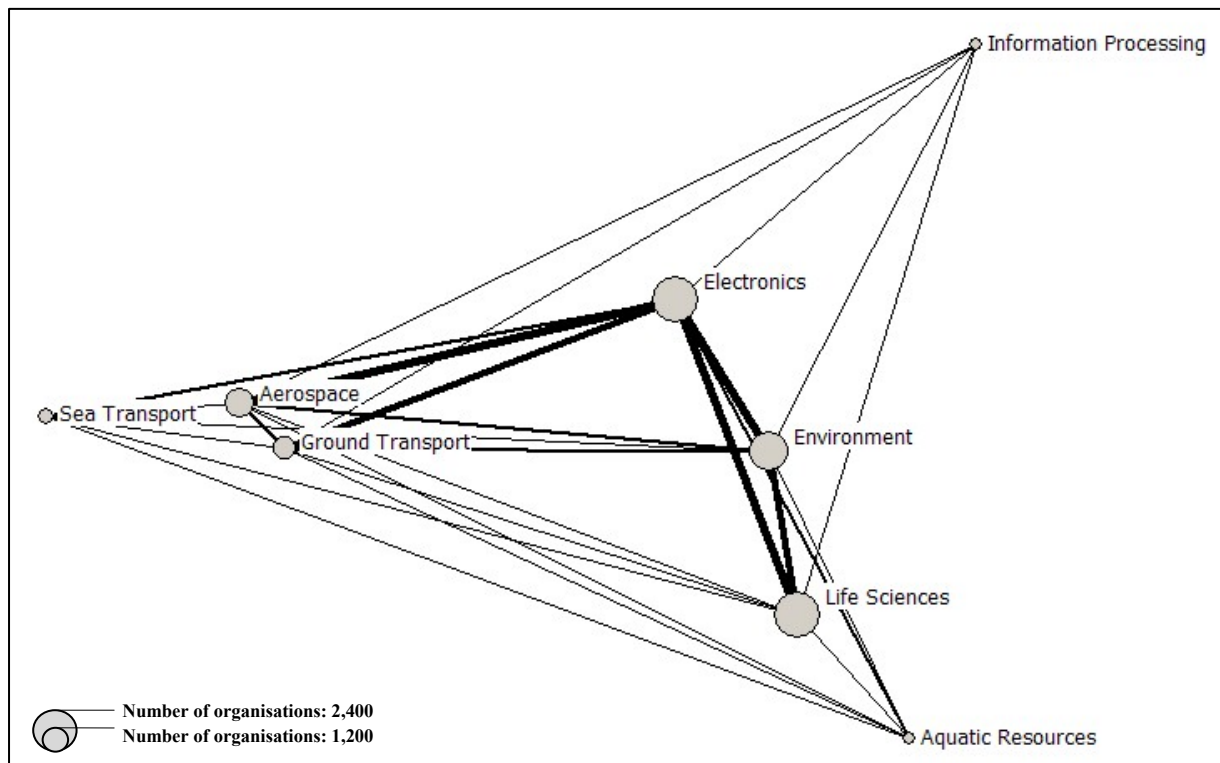
Note: Statistical difference tested using binomial tests whether N_i is different from $E[N_i]$ ***significant at the 0.001 significance level, **significant at the 0.01 significance level, *significant at the 0.05 significance level

The largest communities show the greatest diversity in their substructure. The largest community, *Life Sciences*, shows a broad selection of topics in biotechnology and the life sciences, including health, medicine, food, molecular biology, genetics, ecology, biochemistry, and epidemiology. The second largest, *Electronics*, focuses principally on information technology and electronics, with projects in related fields dealing with materials science, often related to integrated circuits; projects on algorithms, data mining, and mathematics; and a definite subset of projects with atomic, molecular, nuclear, and solid state physics. The third largest community, *Environment*, is focused on environmental topics, including environmental impact, environmental monitoring, environmental protection, and sustainability.

As communities become smaller, they also become more focused. We see, for example, three distinct transportation related communities. The largest of these, *Aerospace*, is focused on aerospace, aeronautics and related topics, including materials science, manufacturing, fluid mechanics, and various energy topics. The next, *Ground Transport* has projects dominated by railroad and, especially, automotive topics; notable subtopics include manufacturing, fuel systems, concrete, and pollution. The smallest transportation community, *Sea Transport*, is more specifically focused; virtually all project titles are shipping-related. The remaining communities, *Aquatic Resources* and *Information Processing*, are the smallest and thematically most uniform.

Figure 3 visualizes the network of key FP5 communities. We determine the position for the communities using methods from spectral graph analysis, so that communities that show a relatively higher number of links between them are positioned nearer to each other (see Seary and Richards 2003). The node size corresponds to the number of organisations of the respective community. It can be seen that the *Life Sciences* and the *Electronics* community show the highest number of organizations. The *Electronics* community appears to have the highest collaboration intensity with other communities, i.e. knowledge produced in this field is used intensively in other fields. The *Life Sciences* community has a strong connection to the third largest community, *Environment*. On the left-hand the three transport related communities are positioned, i.e. they show relatively high inter-community collaboration intensity. The largest of these is *Aerospace*, which is closer to *Ground Transport* than to *Sea Transport*. The community *Aquatic Resources* has the strongest connection to *Environment*, while *Information Processing* is far from all other communities.

Figure 3: Community groups on the network of R&D cooperation



Note: Node positions are determined using methods from spectral graph analysis. Positions are defined as the components of the two most significant eigenvectors, corresponding to the two smallest positive eigenvalues of the normalized Laplacian for the network (See Higham and Kibble 2004 for details).

Structure and Topology of the network communities

Table 2 provides an overview on some measures that characterize the structure of the eight FP5 communities under consideration. When we compare the communities to each other, some noteworthy differences in the network structure appear. As indicated in the previous subsection, the number of vertices, and thus the number of organizations, in a community is highest for the *Life Sciences* community and the *Electronics* community. Though the number of organisations in these two communities is nearly equal, the number of edges is markedly higher in the *Life Sciences* community than in the *Electronics* community, leading to a higher density in the *Life Sciences* community. The average path length also varies across the eight communities. It is highest for the *Environment* community (2.797), though it has a lower number of vertices than the *Life Sciences* and the *Electronics* community, i.e. from a social network analysis perspective, the condition for diffusion of information is better in the latter two communities than in the *Environment* community. The skewness is highest for the *Ground Transport* community showing a value of 6.739. Compared to the other communities,

Ground Transport features central hubs that are in many more projects than the other organizations and are of great importance for the spread of information in the network.

Table 2: Properties of eight FP5 communities

Measure	Life Sciences	Aquatic Resources	Electronics	Environment	Sea Transport	Ground Transport	Aerospace	Information Processing
vertices n	2,366	81	2,307	1,855	218	686	1,146	40
edges m	33,178	451	30,456	23,155	2,978	5,251	13,870	226
Average path length	2.713	2.199	2.732	2.797	2.030	2.549	2.669	1.731
Density	0.012	0.139	0.010	0.013	0.126	0.022	0.021	0.290
Skewness	4.749	1.169	5.132	4.512	1.718	6.739	4.263	1.097
Mean degree	28.046	11.136	26.403	24.965	27.321	15.309	24.206	11.300

Spatial patterns of the network communities

We next consider the spatial distribution of the eight FP5 communities. Figure 4 illustrates the spatial networks of the communities by aggregating individual observations on the organisations of a community to the regional level. The European coverage is achieved by using 225 NUTS-2 regions of the pre-2007 EU25 member states, as well as Norway and Switzerland. Note that the region-by-region community networks are undirected graphs from a network analysis perspective. The nodes represent regions; their size is relative to their degree centrality corresponding to the number of links connected to a region.

The spatial network maps in Figure 4 reveal considerable differences of the spatial collaboration patterns across eight FP5 communities. One important result is that the region Île-de-France takes an important position in all communities. Furthermore, the visualization clearly discloses the different spatial patterns of the Transport related communities, *Aerospace*, *Ground Transport* and *Sea Transport*. Though the region Île-de-France appears to be the central hub in all transport-related communities, the directions of the highest collaboration flows from Île-de-France differ markedly. For the *Sea Transport* community we observe intensive collaborations to important sea ports in the north (Zuid Holland, Agder Rogeland, Denmark, Hamburg) and the south (Liguria, Lisbon, Athens), while for the *Ground Transport* community collaborations to the east and south are dominant (Lombardia, Oberbayern, Stuttgart). In the *Aerospace* community we can observe a strong localisation of collaborations within France and its neighbouring countries.

Figure 4: Spatial patterns of eight FP5 communities

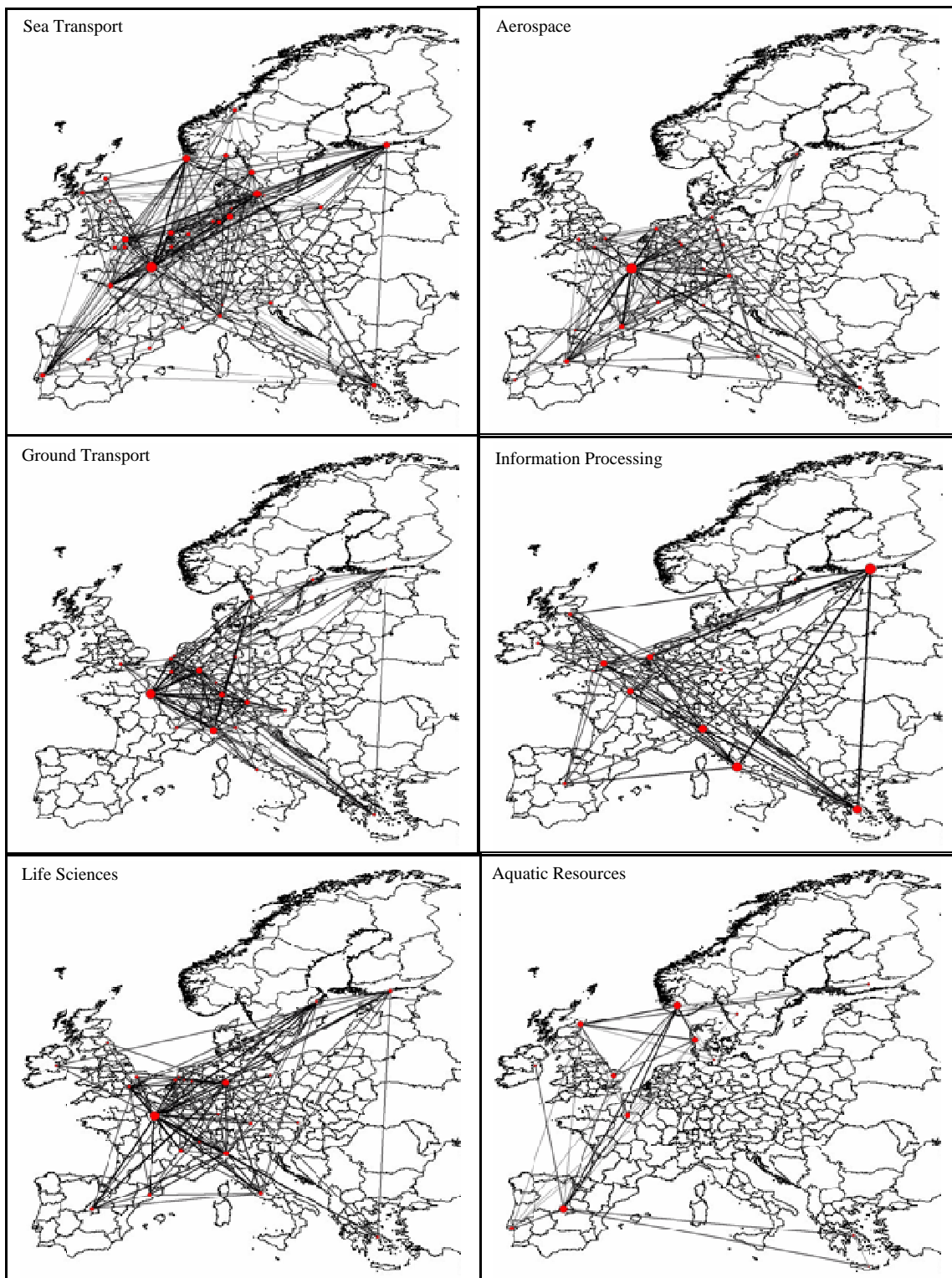
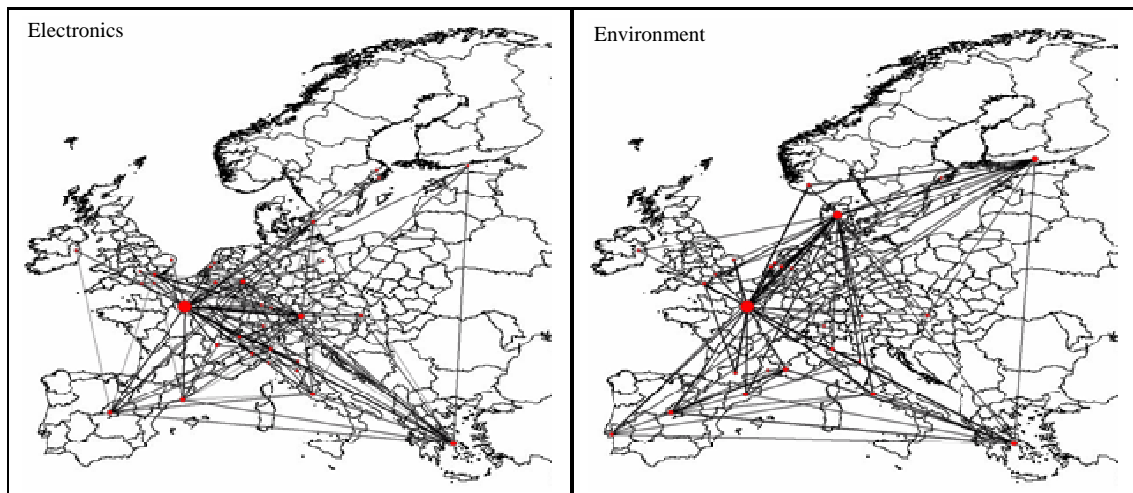


Fig. 4 ctd.



In the largest community, *Life Sciences*, the highest number of collaborations is observed between the regions of Île-de-France and Piemonte (174), while the second largest community, *Electronics*, is characterized by a very high collaboration intensity between the regions of Île-de-France and Oberbayern (474 collaborations), followed by Île-de-France and Köln (265 collaborations), and Oberbayern and Köln (157 collaborations). In the *Environment* community we find the strongest collaboration intensity between Denmark and Helsinki (131 collaborations). In the community *Aquatic Resources* the regions Denmark and Agder Rogaland (Norway) show the highest collaboration intensity, not only between them (21 collaborations) but also to other regions, while for the community *Information Processing* we identify Helsinki as the central region, featuring intensive collaboration with Athens, Lazio and Lombardia.

5 Conclusion

In this paper, we employ recently developed methods to identify communities in European R&D networks using data on joint research projects funded by the European Framework Programmes (FPs). The identification and characterisation of thematically relevant substructures in these networks is of crucial importance in a European policy context. The present study complements earlier empirical work on the structure and topology of R&D networks in Europe that neglected relevant substructures (see, for instance, Breschi and Cusmano 2004).

The network under consideration in the current study consists of nodes representing 25,839 organizations, including firms, universities and research organizations, connected together by joint cooperation in FP5 projects. We adopt a label propagation algorithm (LPA) for identifying community groups in the network under consideration. LPA is designed for maximizing bipartite modularity that accounts for the bipartite character of the network (see Barber and Clark 2009). The advantages of LPA are its conceptual simplicity, ease of implementation and practical efficiency.

The study produces interesting results, both from a scientific point of view, and in a European policy context. We detect eight relevant, thematically relatively homogenous FP5 communities providing a new view on the R&D collaboration landscape in Europe. The larger communities identified are *Life Sciences*, *Electronics*, and *Environment*. However, these may show further relevant substructures. As communities become smaller, they also become more focused. We identified three Transport related communities that are *Aerospace*, *Ground Transport*, and *Sea Transport*. The remaining communities, *Aquatic Resources* and *Information Processing*, are the smallest and most uniform thematically. Furthermore, the results of the spatial analysis clearly reveal that the geographical distribution of the communities varies considerably. However, the region of Île-de-France plays a central role in each of the detected communities.

The general approach followed in this study may be extended and improved upon in several ways. Alternate community detection methods may be considered. More significantly, alternate definitions of what we mean by community may be considered, so as to investigate hierarchical substructures of the communities or to allow the communities to overlap and include the same organization. Other methods from social network analysis may be explored to characterize the network, and techniques from spatial analysis and econometrics may be applied to characterize the network as a whole and its community structure.

Acknowledgements. The authors gratefully acknowledge the grant no P21450 provided by the Austrian Science Fund (FWF).

References

- Barber MJ (2007) Modularity and community detection in bipartite networks, *Physical Review E* 76(6), 066102
- Barber MJ and Clark J (2009) Detecting network communities by propagating labels under constraints, *Physical Review E* 79, 032456
- Breschi S and Cusmano L (2004) Unveiling the texture of a European research area: Emergence of oligarchic networks under EU Framework Programmes, *International Journal of Technology Management. Special Issue on Technology Alliances* 27(8), 747-772
- Caloghirou Y, Vonortas NS and Ioannides S (2002) Science and technology policies towards research joint ventures, *Science and Public Policy* 29 (2), 82-94
- Clauset A, Newman MEJ and Moore C (2004) Finding community structure in very large networks, *Physical Review E* 70, 066111
- CORDIS (2008) List of CORDIS Subject Index Classification Codes, Available from http://cordis.europa.eu/guidance/sic-codes_en.html
- Fischer MM (2001) Innovation, knowledge creation and systems of innovation, *The Annals of Regional Science* 35, 199-216
- Fortunato S (2010) Community detection in graphs, *Physics Reports* 75, 174-189
- Girvan M and Newman MEJ (2002) Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America* 99(12), 7821-7826
- Higham DJ and Kibble M (2004) A unified view of spectral clustering. Mathematics Research Report 02, University of Strathclyde
- Knoke D and Young S (2008) *Social network analysis*. Sage Publications, Los Angeles, London, New Delhi and Singapore
- Leung, IXY, Hui P, Liò P and Crowcroft J (2009) Towards real-time community detection in large networks, *Physical Review E* 79(6), 066107
- Newman MEJ and Girvan M (2004) Finding and evaluating community structure in networks, *Physical Review E* 69, 026113
- Powell WW and Grodal S (2005) Networks of innovators, In Fagerberg J, Mowery DC and Nelson RR (eds.), *The Oxford Handbook of Innovation*, Oxford, Oxford University Press, 56-85
- Raghavan UN Albert R and Kumara S (2007) Near linear time algorithm to detect community structures in large-scale networks, *Physical Review E* 76(3), 036106
- Scherngell T and Barber MJ (2009) Spatial interaction modelling of cross-region R&D collaborations. Empirical evidence from the 5th EU Framework Programme, *Papers in Regional Science* 88, 531-546
- Scherngell T and Barber MJ (2010) Distinct spatial characteristics of industrial and public research collaborations: Evidence from the 5th EU Framework Programme, *The Annals of Regional Science* [forthcoming]
- Searly AJ and Richards WD (2003) Spectral methods for analyzing and visualizing networks: an introduction, In Breiger R, Carley K, and Pattison P (eds), *Dynamic Social Network Modeling and Analysis*, pp 209–228, Washington, D.C. The National Academies Press