The structure of R&D collaboration networks in the European Framework Programmes

Thomas Roediger-Schluga[1,*] and Michael J. Barber[2)]

[1)] Department of Technology Policy, ARC systems research, Donau-City-Strasse 1, A-1220 Vienna, Austria, Phone: +43/(0)50/550-4569, email: thomas.roediger@arcs.ac.at

[2)] Centro de Ciências Matemáticas, Universidade da Madeira, Funchal, Portugal

[*)] corresponding author

This version: 17.08.2006

Abstract

Using a large and novel data source, we study the structure of R&D collaboration networks in the first five EU Framework Programmes (FPs). The networks display properties typical for complex networks, including scale-free degree distributions and the small-world property. Structural features are common across FPs, indicating similar network formation mechanisms despite changes in governance rules. Several findings point towards the existence of a stable core of interlinked actors since the early FPs with integration increasing over time. This core consists mainly of universities and research organisations. We observe assortative mixing by degree of projects, but not by degree of organisations. Unexpectedly, we find only weak association between central projects and project size, suggesting that different types of projects attract different groups of actors. In particular, large projects appear to have included few of the pivotal actors in the networks studied. Central projects only partially mirror funding priorities, indicating field-specific differences in network structures. The paper concludes with an agenda for future research.

JEL Classification: L14, O38, Z13

Keywords: R&D collaboration, EU Framework Programmes, complex networks, small world effect, centrality measures, European Research Area

1. Introduction

Over the past two decades, initiatives to foster collaborative R&D in precompetitive research have become a key instrument of science, technology and innovation (STI) policy at the regional, national and supranational levels. Major international examples are discussed by Caloghirou *et al.* (2002). In Europe, the prime examples are the European Framework Programmes (FPs) on Research and Technological Development (RTD). In these FPs, the European Union has (co-)funded thousands of transnational, collaborative R&D projects; projects aimed at supporting transnational collaboration and coordination in research; and projects supporting transnational mobility for training purposes. Since their inception in 1984, six FPs have been launched and the seventh will commence in 2007.

The main objective of these activities has been to strengthen Europe's science and technology capabilities and to promote European international competitiveness through co-ordinating national policies, integrating national research communities, improving the integration of marginal actors, and bringing together actors with the most advanced resources and capabilities. This has created a pan-European network of actors performing joint R&D.

Strikingly, the composition and structure of this network, in particular at the actor level, has been barely studied to date (notable exceptions are Barber et al. 2006; Breschi and Cusmano 2004). The main reason for this is simply the difficulty of obtaining suitable data – yet the implications are profound.

Although EU-sponsored R&D accounts for only a small part of total R&D in Europe, the FPs are by far the major source of public funding of transnational R&D in Europe. Moreover, research by Larédo and colleagues (see, e.g. Larédo 1998) has shown that, at least in the case of France, most important research actors (large firms, research-intensive small and medium-sized enterprises (SMEs), universities, public research organisations, etc.) participate in EU projects. Therefore, the networks that have emerged in the EU FPs provide valuable information on the organisational fabric and social infrastructure of European science and technology. Knowing how the networks look, how the networks have formed and how the networks evolve in response to external stimuli is of great importance for designing, implementing and assessing new policy measures that aim at creating and deepening the European Research Area.

The networks that have been induced by the EU FPs are very large for social science standards and quite interesting in that they involve a broader set of actors than other sources on

R&D collaborations, in particular data on strategic alliances (for a description of major data sources, see Hagedoorn et al. 2000). Although the networks are moulded by a very particular set of framework conditions and findings thus cannot be generalised naively, the networks provide rich information on a different stage of the innovation process than networks generated from alliance data, patents, scientific publications or surveys.

Using a novel, comprehensive data source on the first five EU FPs – the sysres EUPRO database – we describe key structural features of the networks induced by the Framework programmes and illustrate their economic relevance. This yields insights on patterns of network formation and the social and institutional infrastructure of European research and technology, in particular on what may be the core of the European Research Area.

As formal network formation rules are minimal – a project has to comprise at least two partners from two different countries – we would expect similarly minimal structures. In contrast, however, we observe a great deal of structure that would not be anticipated from the minimal rules.

Prior work (Barber *et al.* 2006) has shown that organisation and project networks in the first four EU FPs are complex networks that share common topological features with many empirical networks in the natural, technological and social domains (see e.g. Strogatz 2001). In particular, they are characterised by a scale-free degree distribution, short characteristic path length and high local clustering – a result that we confirm and extend to FP5 in this work.

Beyond these standard measures, the networks yield additional structural information. Typical patterns of association shed light on network formation and network substructure. Further information is provided by considering the strengths of connections. Projects can be discriminated by project type and central projects can be identified. Organisations can be classified by organisation type and analysed in a similar manner. Further, their participation patterns can be investigated over time. We address each of these points in this paper and identify extant issues to be addressed in future work.

The remainder of the paper is organized as follows. In Section 2, we look at greater depth into the rationale and history of the European Framework Programmes. This is followed in Section 3 by a brief survey of the mathematical structures known as graphs and their use in the analysis of real-world, R&D collaboration networks. In Section 4, we describe the actual networks considered, detailing the raw data on the Framework Programmes, how the data has been refined and regularized, and the procedure by which networks are constructed from the data. In Section 5, we present results on global characteristics, edge properties and vertex properties of networks. In Section 6, we summarise our findings and conclude with an agenda for future work.

2. Rationale and history of the European Framework Programmes

The objectives and instruments of STI policy have changed considerably over the past 20 years, which is also reflected in the evolution of the European Framework Programmes. Institutional framework and objectives have co-evolved, which has impacted the Programmes' designs. At the same time, a number of key elements are common to all FPs. This section briefly highlights the main issues.

The catalytic events for the European Framework Programmes came in the early 1980s in response to widespread concern about the technological competitiveness of European industry, in particular high-tech industries (for excellent accounts of the history of EU research policy, see Peterson and Sharp 1998; Guzetti 1995). In late 1982, the European Commission launched the European Strategic Programme for Information Technology (ESPRIT), upon which the basic structure of the later Framework Programmes was patterned.

ESPRIT provided financial support (about 50% of the project costs) for precompetitive, generic research that had wide applications across many economic sectors. Precompetitive research is located on the continuum between fundamental and applied industrial research, and for practical purposes came to be understood as industrial research that was sufficiently distant from the market to ensure full competition in the product market (for details, see Guzetti 1995, pp. 77-78). This guaranteed compliance with competition rules.

Research in ESPRIT was conducted jointly by firms, research organisations and universities. It involved at least two partners from at least two member states to stimulate transnational linkages and to alert actors to opportunities and needs beyond their home markets (the prior policy of promoting 'national champions' having patently failed). Research results had to be disclosed at least to all members of the consortium. Showing demonstrable success in attracting proposals, several other programmes modelled on ESPRIT were launched shortly afterwards in other strategic technology areas, including telecommunications, industrial technologies, biotechnology and medicine.

At the time, the theoretical rationale for subsidising collaborative, precompetitive R&D was based on the well-known 'market failure' argument. While crucial for long-term competitiveness, precompetitive R&D is an uncertain, risky and increasingly expensive activity, whose results cannot be fully appropriated by any single organisation due to the public good nature of its output. Therefore, subsidies are required to restore private investment incentives and to reap the collective benefits of collaborative R&D in terms of creating critical mass, sharing costs, pooling risk and internalising knowledge spillovers.

In 1984, the various existing and proposed programmes were fused into the First Framework Programme (FP1), institutionalising the model established by ESPRIT. This represented an attempt by the European Commission to design a comprehensive science and technology policy that would give coherence to its RTD efforts and provide a means for selecting European scientific and technological objectives, co-ordinating Community and national policies, and ensuring the necessary funding.

The Single European Act (1987) gave the Community a hitherto missing legal competence for research and technology. It defined the twin objectives of EU RTD policy, 'to strengthen the scientific and technological bases of European industry and to favour the development of its international competitiveness' (Caracostas and Muldur 2001, p. 160). The Maastricht Treaty (1993) added 'cohesion', i.e. narrowing the wealth gap between rich and poor regions, to the objectives of EU RTD policy. Moreover, the Maastricht Treaty stipulated the Commission's power to lead the coordination of national RTD policies, if necessary, and foresaw that all EU activities in the field of research would be included in a multi-annual (five-year) Framework Programme. The latter provision was significant in that it extended the remit of the FPs to include basic research, applied research, technology development and the demonstration of new technologies.

After the termination of FP1 in 1987, the second (1987–1991) and third (1990–1994) remained 'technology-push' programmes in spirit. At about the same time, however, a new theoretical conceptualisation of the innovation process started to pervade policy advisory circles (see, e.g. Soete and Arundel 1993). It conceptualises innovation as a complex, interactive learning process that involves a multitude of actors from all societal spheres (see, e.g. Edquist 2005). The systemic model provides complementary and novel directions for STI policy, including additional rationales for supporting collaborative R&D. These include the need to foster interactive learning as a key mechanisms for knowledge creation; to optimise linkages between the different (sets of) actors involved in innovation processes that rely on increasingly complex knowledge bases; to diffuse new knowledge and technology rapidly and widely; and to build innovative capacity through equipping workers with the requisite knowledge and skills to thrive in an increasingly dynamic, knowledge-based economy (see, e.g. Lundvall and Borrás 2005).

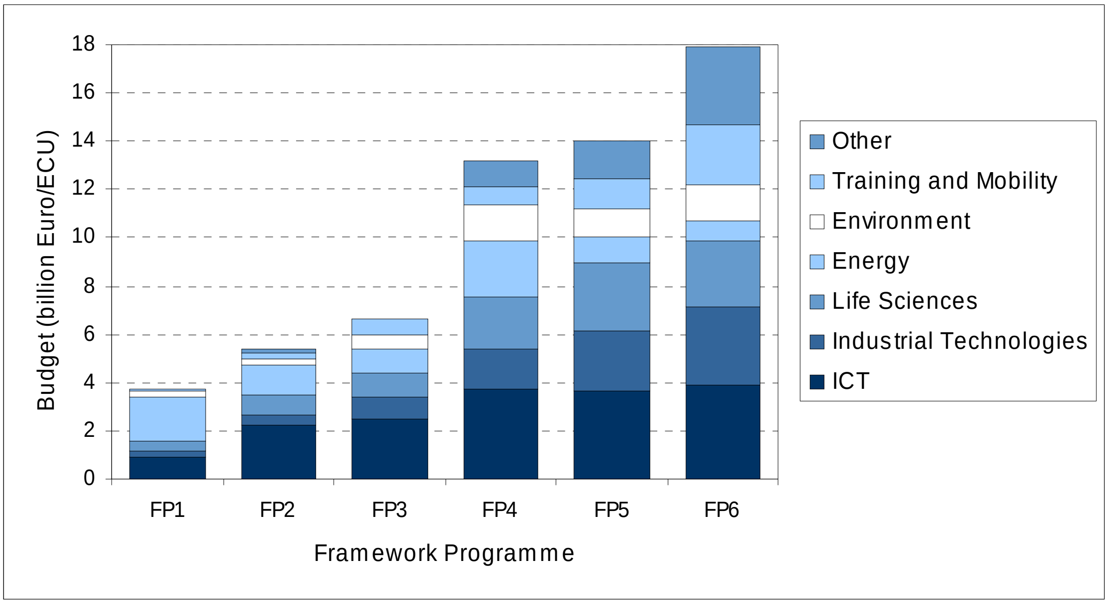
The innovation systems approach has had a profound impact on the orientation of European RTD policy. Much more so than its predecessors, FP4 (1994–1998) focused on the diffusion of new technologies and the integration of SMEs. Moreover, for the first time, it included a substantial budget for training and mobility related measures. Recognising the need for economic and social acceptance of EU-funded research, FP5 (1998–2002) emphasised a stronger user-orientation. It was conceived to help solve problems and to respond to the major socio-economic challenges facing Europe. At the same time, it aimed at maintaining European research capacity and fostering the development of cutting-edge technology.

FP6 (2002–2006) represented a major break with the previous FPs. On the one hand, it focused on scientific and technological excellence in a way that resembled the technology-push-oriented FP2 and FP3. This was done by introducing new instruments – integrated projects and networks of excellence – that brought together a large number of partners. Both were motivated by the perceived need of critical mass of resources and expertise to reach ambitious, fundamental research goals and to create excellent research capacity in Europe. On the other hand, it expanded the scope of the FPs and gave them a new role by becoming the financial instrument to make the European Research Area (ERA) (European Commission 2000) a reality. ERA is intended to overcome the European problems of research fragmentation, underinvestment in R&D, and the lack of co-ordination of national STI policies. This is to be achieved through creating a critical mass and reducing the duplication of efforts by promoting better co-operation and coordination between relevant actors at all levels. FP7, which commences in 2007, will continue in this direction, deepening ERA and carrying it further towards the development of the knowledge economy and society in Europe (CORDIS 2006a). It is designed for a period of seven years, until 2013.

Figure 1 shows that available research funding in the FPs has grown considerably over time from € 4bn in FP1 to almost € 18bn in FP6. The largest increases occurred from FP3 to FP4 and F5 to FP6. Breaking down total funding by the main thematic priorities shows that the largest share of research funding has gone to information and communication technologies (ICT) in all but the first Framework Programme. Research funding for industrial technologies and life sciences (including food and agriculture) has increased markedly over time, both in absolute and in relative terms. Also, the environmental thematic priority (including transport) has become quite sizeable in FP4 to FP6. In contrast, energy, which accounted for almost half of total funding in FP1, has become considerably less important in recent FPs. Absolute funding for energy is actually smaller in FP5 and FP6 than in FP1.

Within these thematic priorities, the main funded activities are shared cost research and the co-ordination of national research activities that may range from fundamental basic research to fairly applied development work and validation with users. Since FP4, R&D activities have been complemented with funding for the training and mobility of researchers, special support for SMEs, networking and exploitation activities. This is reflected in Figure 1 in the growing share of the categories 'training and mobility' and 'other'.

Figure 1: Budget FP1–FP6: Evolution and share of thematic priorities



Note: ICT ... information and communication technologies; industrial technologies include materials, aeronautics and space technologies; life sciences include biotechnology, genomics, biomedicine and food; environment includes transport; Other includes support for SMEs, dissemination, demonstration, co-operation with third countries, and ERA related measures.

Source: adapted from CORDIS (2006c; 2006b); European Commission (2006); Barker and Cameron (2004, p. 172).

Despite the evolution of their objectives and their scope (an excellent comprehensive yet concise account is Barker and Cameron 2004), the fundamental rationale of the FPs as mid-term research programmes that support collaborative research in selected technological priority areas has remained unchanged. Moreover, all FPs share a few key structural elements (see Caracostas and Muldur 2001, p. 162). In particular,

- the EU only co-funds projects of limited duration that mobilise private and public funds at the national level,
- the focus is on multinational and multi-actor co-operations that add value by operating at the European level (see European Commission 2002),

- all projects are proposed by self-organised consortia, and
- the selection for funding is based on specific scientific excellence and socio-economic relevance criteria.

Because of these common features, projects and networks can be compared over time. The remainder of the paper is devoted to this task.

3. Formal and conceptual background

Before examining the network data, we need to establish the formal and theoretical concepts relied upon in this work. This is done in the present section.

3.1. Graph concepts

A network is a somewhat loose term describing an object that is composed of elements and connections between these elements. The concept can be made rigorous by identifying networks with graphs. A graph is a mathematical structure consisting of a set of vertices (nodes) and a set of edges (links) that connect pairs of vertices. Graphs are a versatile model class that have become increasingly popular in recent years as a basis for investigating natural, technological and social networks (for an overview, see Bornholdt and Schuster 2003).

A graph can be represented by a matrix **A** called the *adjacency matrix*. The adjacency matrix has elements A_{ij} equal to 1 if there is an edge between vertices i and j and equal to 0 otherwise. If the network is *weighted*, e.g. some edges represent stronger connections than others, elements A_{ij} may be generalised to take on real number values to represent stronger and weaker connections. If **A** is a symmetric matrix, the graph is *undirected*, otherwise it is *directed*.

The total number of edges connected to a vertex (total number of adjacent vertices or nearest neighbours) is the *degree k*. In directed graphs, the number of incoming and outgoing edges are separately totalled to produce the in-degree and the out-degree. The *density* of a graph is defined as the ratio of the number of edges present in the graph to the number of edges that could be present.

Graphs can be partitioned into subgraphs, i.e. subsets of vertices and edges. The simplest is the dyad, consisting of two vertices that are either adjacent, i.e. connected by an edge, or not. Subgraphs of size three are called triads. If the three vertices are connected, they constitute a *triangle*. A triangle is the smallest nontrivial example of a *clique*. Cliques are *complete* (or fully connected) *subgraphs*, in which every vertex is adjacent to every other. Cliques have the property that *transitivity* holds within the clique; i.e. if vertex i is a neighbour of vertex j and j is a neighbour of vertex k, k is also a neighbour of i.

A graph is *connected* if it is possible to establish a path from any vertex to any other vertex of the graph. A *path* is the alternating sequence of vertices and edges, starting and ending with a vertex, in which each vertex is incident with the lines following and preceding it in the sequence.

Paths are useful to measure *distance*, i.e. how far apart vertices are in a graph. The *shortest path* between two vertices is referred to as a *geodesic*. The average geodesic in a connected graph is the *characteristic path length* ℓ. The maximum geodesic from vertex i to any other vertex is its eccentricity. The maximum eccentricity in a graph is its *diameter*.

An empirically important special class are bipartite graphs, or affiliation networks, where vertices can be partitioned into two disjoint subsets, with edges existing only between the two sets. These are appropriate models for many social networks, where ties between actors are defined by their joint affiliation to organisations, events, etc. The networks we consider in this paper are bipartite graphs, with the two sets comprising organisations and projects.

3.2. Graph models

The simplest baseline model for empirical networks is a random graph. In random graphs, graph properties such as the number of vertices, the number of edges, or the connections between them are determined in some random way. Random graphs are useful for the mathematical determination of expected properties of graphs (see Bollobás 2001). For empirical applications, random graphs provide a baseline for comparison between real-world networks and networks generated by putative network formation rules. Thus, a random graph model can play the role of a theory to be tested by comparison to real-world networks.

An important class of random graphs are Erdős-Rényi random graphs (often simply called "random graphs" due to their historical role). In the Erdős-Rényi model, an edge exists between any pair of vertices with probability p. Many properties of this simple model are exactly solvable in the limit of large graph size. In particular, such graphs undergo a phase transition as p increases, transforming from multiple disconnected components to a single *giant component* to which all but a negligibly small fraction of the vertices are connected. Random graphs have short characteristic path lengths. This is also observed in real networks, where it is known as the *small world* effect, and allows each vertex in the giant component to reach any other vertex in a small number of steps.

However, Erdős-Rényi random graphs lack many other properties observed in real networks. In particular, the diameter grows logarithmically with the number of vertices; local clustering, describable in terms of the number and size of cliques, is low; the distribution of degrees has a Poisson form; and the graphs show no correlation between the degrees of adjacent vertices. In light of this, the Erdős-Rényi random graph model can be viewed as only a starting point for investigating real networks.

A seminal model that reproduces the small world effect is due to Watts and Strogatz (1998). In their *small world network* model, they begin with a locally connected graph and randomly rewire connections, giving rise to 'shortcuts' outside the local region. This change produces networks that show the small world effect and feature the high clustering seen in many real networks.

Further refinements yield models that feature additional properties missing from the Erdős-Rényi model and the Watts and Strogatz model. In this work, we are particularly interested in *scale-free* network models. Scale-free networks are characterized by having a power-law degree distribution, and can feature the formation of a giant component, the small world effect, substantial local clustering, and correlations between the degrees of neighbouring vertices.

3.3. Relating graphs and networks

3.3.1. Global topology

The global topology of a complex network (for reviews, see Dorogovtsev and Mendes 2004, 2002; Newman 2003a; Albert and Barabasi 2002) is usually recognised through three indicators: a scale-free or power-law degree distribution, a short characteristic path-length, and high average clustering. These properties have also been identified in R&D collaboration networks, such as in the life sciences (Powell et al. 2005) and in the networks stimulated by the European Union in its Frameworks Programmes (Barber et al. 2006; Breschi and Cusmano 2004). We will briefly describe each of these indicators and highlight their empirical implications.

In classical random graphs, the degree distribution, given by the average fraction of vertices of degree k, has a Poisson form (see, e.g Bollobás 2001). The degree distribution of many large real-world networks, however, in such distinct domains as the technological, biological, psychological and social realms (see Newman 2003a), is strongly right-skewed with a heavy tail. Mathematically, the tail can be described by a power law (Barabasi and Albert 1999),

$$p(k) \sim ck^{-\alpha},$$

where c and α are positive constants. This means that the probability p of a vertex selected at random having a degree k declines with the size of α determining the rate of decay. The size of α also determines the convergence or divergence of the higher moments of the empirical degree distributions (see technical appendix). Such divergences can produce strong effects, such as high degree vertices with noticeable probability (see Dorogovtsev and Mendes 2004).

The power law can be identified by plotting the degree distribution on logarithmic axes, where it appears as a straight line with slope $-\alpha$. This contrasts with a normal or a Poisson distribution which drops off sharply on a log-log plot, such that the probability of a degree greater than a cut-off value is effectively zero (for a nice graphical illustration, see Watts 2004, p. 251). A cut-off thus implies a characteristic scale for the degree distribution of a network. Since a power-law degree-distribution lacks any such cut-off value, it is often called a scale-free distribution and the so-characterised network scale-free.

This property has important empirical implications. While the majority of vertices have a less-than-average degree, some vertices have a degree that is orders of magnitude larger than the average. These 'hubs' are centrally located and highly interconnected vertices that may dramatically affect the way a network operates.

3.3.2. Knowledge creation and knowledge diffusion

A necessary condition for information or objects to spread or to diffuse in a network is that vertices are connected, either directly or indirectly through a path of connected vertices. In a large scale-free network, most vertices form a giant component such that they are all mutually reachable. This giant component is highly resilient to the random removal of vertices, as most vertices are only linked to a few others (a property referred to as 'attack tolerance' or 'ultraresilience'). However, if the highly connected hubs are removed, the giant component quickly falls apart into smaller, disconnected components, disrupting any global transmission process.

Another property of the giant component in scale-free networks is that it has a short diameter, i.e. the maximum number of steps between two vertices is small. At the same time, scale-free networks are quite dense locally: nodes tend to be connected with many of their direct neighbours. A quantity of particular interest are triangles, i.e. triples of connected vertices.

Dividing the number of actual triangles of which vertex i is a part by the number of theoretically possible triangles (if all neighbours were connected) yields the clustering coefficient C_i for vertex i. Averaging C_i over all vertices gives the network-specific clustering coefficient C. In scale free networks this is considerably higher than in random networks, a property that is also known as *cliquishness*.

Networks that combine a short average path length with a high clustering coefficient are called *small-world networks* (Watts and Strogatz 1998). These are quite interesting in terms of knowledge creation and knowledge diffusion, two key functions of R&D collaboration networks.

When path lengths are short, new knowledge can spread rapidly and widely through the population and thus fuel local knowledge creation. This is a necessary but not sufficient condition for learning. Like in a barter economy, there has to be a double coincidence of wants for knowledge exchange to take place. This becomes clear if we bear in mind that completely identical agents cannot learn from each other. Rather, agents have to know different things while being sufficiently similar to be able to communicate and share complementary knowledge. Communicating through joint neighbours relaxes this constraint. Such local redundancy is present in cliquish networks but absent in Erdős-Rényi random graphs.

Knowledge transmission is limited by absorptive capacity, the ability to make sense of and to leverage new knowledge. Knowledge thus degrades as it is passed along long chains, which is costly in terms of time and the diminution of knowledge. If new knowledge is difficult to absorb (e.g. because a considerable part is tacit) or if transmission requires repeated interaction, the redundancy of ties in cliquish networks is again beneficial, as it facilitates the validation of new knowledge and the possibility for multiple interactions. Note that the value of cliquishness critically depends on absorptive capacity. If information transmission is (near) perfect, nothing can be learned afterwards and there is no benefit in redundancy. Instead, links between cliques are more beneficial.

Cowan and Jonard (2004; 2003) formalise these ideas. They show that small world networks generated by the Watts and Strogatz (1998) model initially produce relatively rapid aggregate knowledge growth in settings where absorptive capacity is low. Long-run knowledge levels are higher in small world networks than in other network configurations. Thus, in exploration networks (Rothaermel and Deeds 2004), such as pre-competitive R&D collaboration networks, where knowledge is less-codified and there is a great deal of diversity in search activities, the optimal network structure for communication and knowledge flow is cliquish. In contrast, in exploitation networks (Rothaermel and Deeds 2004) that aim at exploiting existing capabilities, optimal network structures are more random (Cowan et al. 2004).

3.3.3. Mixing patterns and network formation

So far, we have treated all vertices identically, which is seldom the case for real-world networks. Rather, vertices may differ in terms of numerical or categorical properties, e.g. weight or organisation type. Taking this additional information into account can provide considerably more information on network structure and also shed some light on network formation rules. This is closely related to the notion of homophily, which is the tendency of actors to associate and bond with similar others (see McPherson et al. 2001).

A straightforward and much studied measure of homophily is based on the degree. If actors tend to associate with others who are like them, i.e. neighbouring vertices tend to have similar degrees, the network is said to show *assortative mixing* or assortative matching. If actors prefer to associate with others who are different, the network shows *disassortative mixing* (Newman 2003b, 2002).

The nature of the mixing has been measured in a variety of ways. Newman has proposed a simple and flexible measure based on the use of the Pearson correlation coefficient between degrees of neighbouring vertices. By examining the correlation coefficients for several real world networks of diverse origins, he concludes that social networks tend to be assortatively mixed, while biological and technological networks tend to be dissortatively mixed.

Assortative mixing can have profound effects on the structural properties of a network. If mixing occurs by some discrete attribute, the network will tend to break into separate communities, which may impact diffusion processes. Likewise, Newman shows that in networks with positive degree correlation, a giant component forms more easily and is more closely connected than in networks with no or negative degree correlation. The higher density of assortatively mixed networks suggests faster diffusion. At the same time, the giant component is more resilient to the random removal of central vertices, presumably because these cluster together in the core group and are therefore to some extent redundant. While these properties are undesirable in the context of, e.g., epidemic spread of diseases, they are quite desirable for information and knowledge diffusion.

3.3.4. Centrality measures

One of the fundamental properties of a node is its position in the network. In mathematical sociology, several measures have been developed to quantify vertex *centrality* (see Wasserman and Faust 1994, and the references cited therein) that allow to identify and rank vertices by their structural importance. The assumption is that central actors have a stronger influence on other network members. We will consider four centrality measures, each of which quantifies a different aspect of centrality. Conveniently, each of them can be normalised to the interval [0,1] and can be compared across networks of different size.

The simplest measure is the *degree centrality*, which is defined as the ratio of degree k_i and the maximum degree k in a network of the same size. It is based on the idea that vertices that are connected to many neighbours have power in many social settings. They are highly visible and should be recognised by others as major channels of relational information.

A more sophisticated version of the same idea is known as *eigenvector centrality*. Having many connections surely affords influence and power, but not all connections are the same. Typically, connections to actors who are themselves well connected (high degree) will provide actors with more influence than connections to poorly connected (low degree) actors. Eigenvector centrality thus accords each vertex a centrality that depends both on the number and the quality of its connections by examining all vertices in parallel and assigning centrality weights that correspond to the average centrality of all neighbours. This is done using the adjacency matrix representation of the graph and produces a vector of centralities that is an eigenvector of the matrix corresponding to the largest eigenvalue (see technical appendix). The eigenvector centrality is mathematically very natural because it is based on fundamental properties of the adjacency matrix, specifically, the spectral properties of the matrix.

Another way to define centrality is based on network paths. Assuming that information, infections, commodities, etc. take the shortest paths when spreading in a network, vertices that are at a short distance from any other are likely to receive them more quickly than more distant vertices. This idea is quantified by the *closeness centrality* of vertex i, which is defined as the inverse of the mean geodesic distance (i.e., the mean length of the shortest path) from vertex i to every other vertex in a connected graph.

Based on the same logic, the *betweenness centrality* of vertex i can be defined in a connected graph as the fraction of geodesic paths between any pair of vertices on which i lies. Betweenness centrality is commonly interpreted as a measure of control, as actors that lie on many shortest paths can exert considerable control and act as gatekeepers.

4. Framework Programme networks

4.1. The data set

We draw on the latest version of the sysres EUPRO database. This database includes all information publicly available through the CORDIS projects database (CORDIS search 2006) and is maintained by ARC systems research (ARC sys). For purposes of network analyses, the main obstacle is the inconsistency of the raw data. Apart from incoherent spellings in up to four languages per country, organisations are labelled inhomogeneously. Entries may range from large corporate groupings, such as EADS, Siemens and Philips, or large public research organisations, such as CNR, CNRS and CSIC, to individual departments and labs.

These are listed as valid at the time the respective project was carried out. Among heterogeneous organisations, only a subset contains information on the unit actually participating or on geographical location. Information on older entries and the substructure of firms tends to be less complete. These properties rule out any fully automated standardisation method.

In order to homogenise the data, organisational boundaries are defined by legal control and entries are assigned to the respective organisations. Resulting heterogeneous organisations, such as universities, large research centres, or conglomerate firms are then broken down into subentities that operate in fairly coherent activity areas, such as faculties, institutes, divisions or subsidiaries. Based on the available contact information of participants, subentities have been identified for a significant number of entries.

This procedure reflects the two main approaches in the modern theory of the firm, which define organisations as a) a cost-minimising contractual agreement or as b) a (coherent) bundle of resources and competencies (see, e.g. Hodgson 1998). Creating roughly comparable units across organisations mitigates the fundamental problem of the appropriate scale at which organisations should be compared. Moreover, coherent organisations yield more precise information on actual collaboration patterns.

The case of the French Centre National de la Recherche Scientifique (CNRS), the most active participant in the EU FPs, may serve as an illustration. Initially, 2737 separate entries were summarized under a unique organisational label. These 2737 entries were broken down into the eight areas of research activity in which CNRS is currently organized. Based on available information on participating units and geographical location, 2650 of the 2737 entries could be assigned to one of the eight subentities. For the remaining 87 entries, the nonspecific label CNRS was used.

Comparable processing was done for other large research organisations and universities. However, due to scarcer information, firms could not be broken down into subentities at a comparable rate. Moreover, owing to resource constraints, standardization work has focused on the major players in the FPs. Organisations participating in fewer than a total of 30 projects in FP1–FP5 have not been broken down yet.

Wherever possible, missing information is added and the resulting dataset is regionalised according to the European NUTS (Nomenclature of territorial units for statistics) classification system (European Commission 2005). Particular attention is devoted to cleaning the poor-quality raw data on organisation types. To the best of our knowledge, this makes the sysres EUPRO database the most complete and highest quality data source on the European Framework Programmes[1].

Table 1 shows that the sysres EUPRO database presently comprises information on 43,317 projects over the period 1984 (first project starting dates) to 2012 (last scheduled project end date). At its present state of standardisation, the database includes 42,020 separate organisations that were involved in at least one project. This figure increases to 49,885 if we consider subentities. Data on the first four FPs is complete according to the CORDIS website. In FP5, the database is missing 554 projects which have been added to the CORDIS project database since the latest update of the sysres EUPRO database. The database will be updated with the missing data and information on FP6 as it becomes available.

Participating organisations are either coded as prime contractor (i.e. co-ordinator) or participant. There is no further information on participants' roles. Information on prime contractors is available for virtually all projects. Although projects by definition have to comprise at least two partners, information on additional participants is only available for a subset (see Table 1). However, this subset comprises a sizeable majority of the population for all FPs beginning with the second. This also applies to FP4 and FP5, where the apparent decline in projects with information on multiple partners is due to the addition of training, mobility and supportive measures, which mostly list only the main applicant.

---

1 We are only aware of one comparable major data source on the EU FPs, the EU RJV database, which is part of the STEP-TO-RJVs database (Caloghirou and Vonortas 2000). It has been constructed in the TSER project 'Science and Technology Policies Towards Research Joint Ventures' and contains information on all projects funded in FP1-FP4 that have at least one participant from the private sector. The 6,300 research joint ventures, however, represent only a subset of the corresponding 20,700 projects with information on more than one participant included in the sysres EUPRO database.

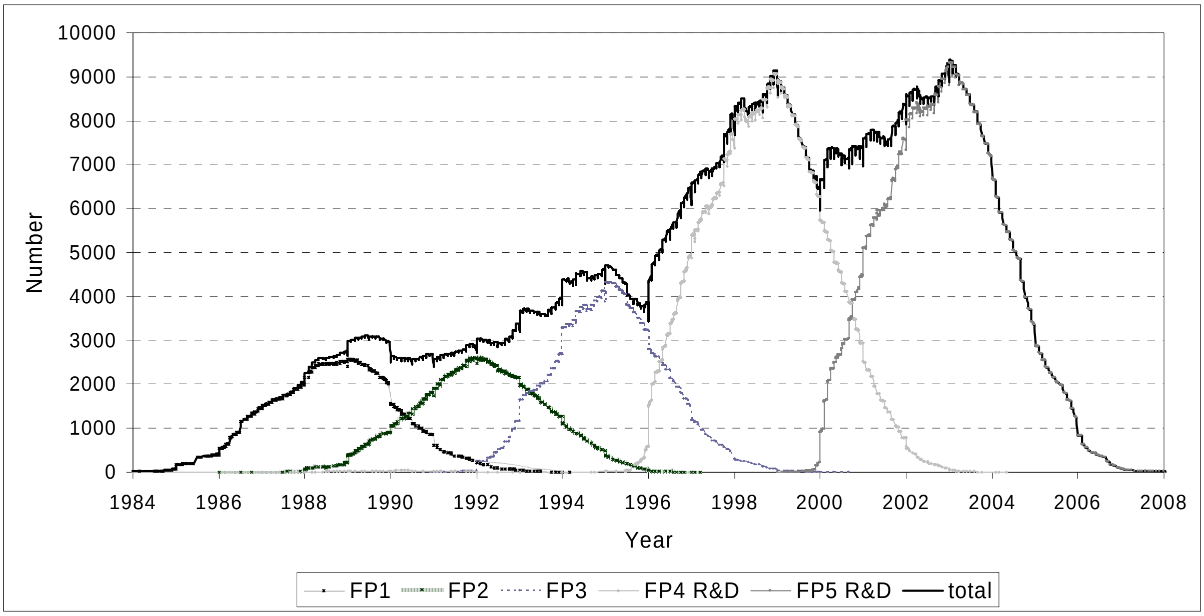
Table 1: sysres EUPRO database – numbers of projects and organisations described

Framework Programme (FP)	Period	Projects	Projects with multiple partners	Organisations	Subentities
FP1	1984–1987	3,283	1,696	1,981	2,583
FP2	1987–1991	3,885	3,013	4,572	6,300
FP3	1990–1994	5,529	4,611	7,324	10,025
FP4	1994–1998	15,061	11,374	19,755	24,156
FP5	1998–2002	15,559	10,674	22,303	27,382
Total		43,317	31,345	42,020	49,855

Note: EURATOM projects are not listed. Recipients of research grants are not counted as organisations or subentities.

By examining the number of active R&D projects, the above figures can be related to actual research activity over time. Figure 2 shows that the total number of active R&D projects increases until FP4 and reaches about the same level in FP5. Moreover, the figure makes clear that there is considerable overlap between the different FPs that is not only caused by the temporal overlap between them. According to the available data, it takes up to two years after the official beginning of an FP (FP3-FP5) for a sizeable number of R&D projects to kick-off. As the average duration of R&D projects is in the range of 31–35 months in each FP, the number of active projects peaks past the official end date of each FP. Projects funded in the final calls of the FPs may last more than four years past the FPs' official termination.

Figure 2: Number of R&D projects over time, FP1–FP5



Note: Only indirect actions, i.e. research carried out by third parties and co-funded by the EU, are considered. In the data on FP4 and FP5, non R&D projects are excluded. These are preparatory, accompanying and support measures (ACM), exploratory awards (EAW), Access to Research Infrastructures (LFC), Research network contracts (NET), Research grants (individual fellowships) (RGI), Bursaries, grants, fellowships (BUR), and Research Infrastructure-Transnational access (TA).

Although the sysres EUPRO database already incorporates massive efforts to produce a reliable data source on activities funded in the EU FPs, there is scope for improvement. Apart from continuously updating the database and improving the homogeneity of the data, the main challenge is to make the data consistent over time. At the moment, participants are labelled with their most recent name. This ensures continuity but introduces a bias in that many legally separate participants have become parts of new organisations at a later stage. Also, a sizeable number of organisations have only been founded recently, such as EADS or Helmholtz-Gemeinschaft. Tracing the genealogy of actors, in particular of firms, has only started recently. This will provide the data required for detailed dynamic analyses in the future.

4.2. Network construction

In our empirical analyses, we focus on organisation and project networks separately, rather than analysing them as bipartite networks. Since we do not have detailed information on the intra-project structure, we have to construct the networks from our data. This we do by projecting the bipartite graph onto the set of organisations and the set of projects, producing the O-graph and P-graph, respectively. We assume that organisations (projects) are connected if they share a project (an organisation). In other words, if Alcatel and ABB participate in the same project, an edge is drawn in the O-graph. If Shell participates in two projects, an edge is drawn in the P-graph. Edges can be weighted with the cardinality of the intersection, e.g. if Alcatel and ABB participate jointly in two projects, the corresponding edge has a weight of two and so on. The *size* of each vertex is its degree in the bipartite graph, e.g. a project comprising ten organisations has size ten, as does an organisation participating in ten projects.

In constructing the networks, we thus assume each project to be a fully connected subgraph of organisations, i.e. a clique, and similarly for organisations and the projects they participate in. This is an idealised graph type that, although not fully representative, is a reasonable approximation to the actual intra-project structure of all but very large projects. Since the vast majority of projects in our data set have fewer than 15 participants, our construction rule is considerably more accurate than assuming the other idealised type of a star structure, in which each participant is only connected to the project coordinator as central vertex.

To keep our analyses consistent across the FPs, we only consider R&D projects and exclude all training, mobility and accompanying measures in FP4 and FP5. We construct the networks using information on subentities, if available, since this yields more accurate information on actual collaboration patterns. As this information is not consistently available for all organisations, we cannot rule out the possibility that the process of breaking down organisations into subentities has introduced a bias into our data analysis. However, we have run all the reported analyses during the continual refinement of the data and have obtained qualitatively and quantitatively similar results, apart from extreme values, e.g. the maximum degree.

5. Empirical characteristics

5.1. Global characteristics

In Table 2a and Table 2b, we give some basic properties of the project and organisation networks for the FP1–FP5. The facts that FP1 was the first program launched and that the available data are rather incomplete make it exceptional in many respects. We therefore focus our analyses on FP2–FP5, providing graph characteristic values for FP1 merely to indicate the difference from the networks created by the subsequent FPs.

Table 2a: Basic network properties of FP1–FP5 organisation projection

Graph characteristic	FP1	FP2	FP3	FP4	FP5
No. of vertices N	2,612	6,397	10,158	22,501	23,116
No. of edges M	9,645	64,459	114,826	244,136	324,361
No. of components	413	217	453	467	83
N for largest component	2,102	6,084	9,460	21,775	22,788
Share of total (%)	80.5	95.1	93.1	96.8	98.6
M for largest component	9,495	64,282	114,367	243,418	323,653
Share of total (%)	98.4	99.7	99.6	99.7	99.8
N for 2nd largest component	8	6	9	45	12
M for 2nd largest component	22	15	36	10	48
Clustering coefficient	0.563	0.717	0.717	0.787	0.806
Diameter of largest component	10	7	8	10	11
ℓ largest component	3.77	3.34	3.39	3.54	3.36
Mean degree	7.39	20.15	22.61	21.7	28.06
Fraction of N above the mean (%)	28.9	28.4	25.2	23.7	24.6
Mean vertex size	3.06	3.02	3.11	2.75	2.62
Standard deviation	5.47	5.17	6.19	5.92	5.63

Examination of the numbers of vertices and edges in the connected graph components indicates that a giant component is present in all Framework Programme networks. In each case, the great majority of vertices and essentially all edges are in the giant component, supporting the choice to focus on the giant component.

As expected, all networks are of small world type: They exhibit a high clustering coefficient and a small characteristic path length. This is a positive result in terms of what we presently know about knowledge creation and knowledge diffusion in exploration networks (see Cowan 2006).

Table 2b: Basic network properties of FP1–FP5 project projection

Graph characteristic	FP1	FP2	FP3	FP4	FP5
No. of vertices N	3,283	3,884	5,529	9,471	6,803
No. of edges M	42,804	87,956	186,039	386,205	339,881
No. of components	413	217	453	467	83
N for largest component	2,709	3,626	5,031	8,948	6,714
Share of total (%)	82.5	93.4	91.0	94.5	98.7
M for largest component	42,497	87,898	185,990	386,129	339,873
Share of total (%)	99.3	99.9	100.0	100.0	100.0
N for 2nd largest component	10	6	3	9	3
M for 2nd largest component	45	15	3	38	3
Clustering coefficient	0.654	0.541	0.448	0.479	0.438
Diameter of largest component	9	7	9	9	10
ℓ largest component	3.36	2.86	2.80	2.79	2.55
Mean degree	26.08	45.29	67.3	81.56	99.92
Fraction of N above the mean (%)	36.2	37.0	35.5	36.2	35.8
Mean vertex size	3.74	6.04	6.61	7.23	8.94
Standard deviation	2.14	4.59	4.59	4.63	6.07

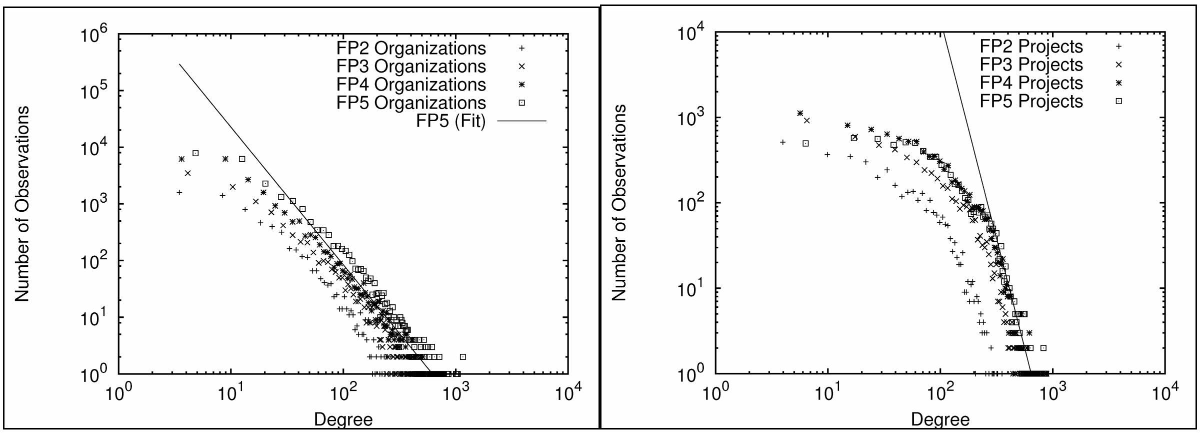
There is a slight increase in the clustering coefficient of the organisation networks from FP1 to FP5. This suggests that integration between collaborating organisations has increased over time, indicating that Europe has already been moving toward a more closely integrated European Research Area in the earlier Framework Programs.

Conversely, there is a slight decrease in the clustering coefficient amongst the projects. Taken together, these trends appear to indicate changes in the interaction patterns as the average project size increases with later FPs, while the average number of projects per organisation remains roughly constant across FPs.

Qualitatively, the changes are not inconsistent with smaller projects being grouped together into larger projects in name but not in fact to meet changing funding expectations. Such groupings would tend to increase the number of transitive connections (i.e. triangles) in the O-graphs and to decrease the number of transitive connections in the P-graphs, with corresponding effects on the clustering coefficients. However, this pessimistic interpretation would depend on both bad faith in project applications and poor quality in application evaluations. Although possible in some cases, this seems unlikely as the sole explanation of the observed trends. A more complete explanation of the trend could be provided by a closer examination of the interactions of organisations within projects – for which we lack suitable data at present – or, possibly, inferred from a study of dynamic changes in network structures throughout the FPs.

The mean degree in the P-graphs shows a marked increase with time. This is contrasted by the mean degree in the O-graphs, which shows no pronounced change over time. This can be interpreted as organisations having a roughly constant capability to maintain connections to one another, while projects have no such constraint. A greater number of participants per projects increases the total ability to make connections, which is consistent with the observed rise of the mean degree in the P-graphs.

Figure 3: Degree distribution of FP2-FP5 organisation and project projections



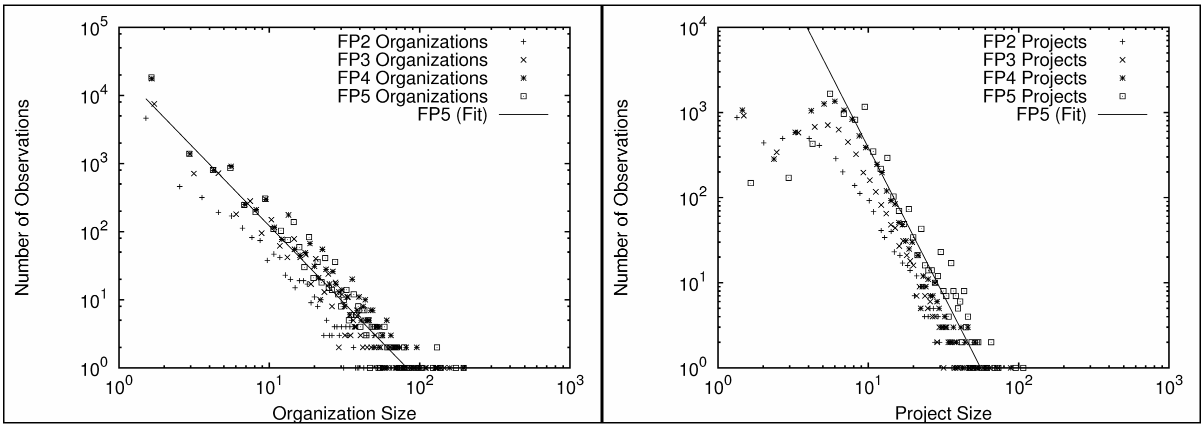
Note: Exponent of the fit line in the organisation projection is $\alpha = 2.43 \pm 0.07$. Exponent of the fit line in the project projection is $\alpha = 3.1 \pm 0.2$. Standard error is calculated using bootstrapping with 1000 samples. Fit lines for earlier FPs are similar (not shown).

In both projections, the distributions of degrees and vertex sizes are highly skewed. This is seen in the fraction of vertices with degree above the mean, and in the relative sizes of the mean vertex size and the standard deviation of the same. In fact, the distributions for both projections are scale-free in the right tail (for details on how power laws are fit, see Technical Appendix). The degree distribution for the organisation network is scale-free beginning at a degree of about 10 (see Figure 3). In contrast, the degree distribution for the P-graph has a more complicated structure with a scale-free character only at the extreme end of the tail. Theoretical arguments (see Barber et al. 2006) suggest the interpretation of the degree distributions for the P-graphs as the combination of two distinct power laws. However, fitting power laws to the actual data does not produce a high-quality fit, especially in earlier FPs; further investigation appears warranted.

The scale-free structure of the degree distributions has profound consequences, essentially stemming from the spread of degrees over several orders or magnitude. Perhaps the most apparent such consequence is the presence of network hubs: these high degree vertices are the most visible members and tend to be pivotal for the coherence of the network. The scale-free distribution indicates a great deal of heterogeneity amongst the network participants that must be taken into account when investigating the nature of the networks.

Similarly, the vertex size distributions have scale-free characteristics. The distribution of organisation sizes is scale-free for essentially all values (see Figure 4). It is remarkably similar across FPs, indicating that the distribution of organisations able to carry out a particular number of projects has not changed over time. A complementary interpretation of this finding is that the changes in the underlying research activities have not altered the mix of organisations participating in a particular number of projects in each FP.

Figure 4: Distribution of vertex sizes in the FP2-FP5 organisation and project projections



Note: Exponent of the fit line in the organisation projection is $\alpha = 2.28 \pm 0.05$. Exponent of the fit line in the project projection is $\alpha = 3.5 \pm 0.2$. Standard error is calculated using bootstrapping with 1000 samples. Fit lines for earlier FPs are similar (not shown).

The distribution of project sizes, as measured by participation, is indicative of a typical range of between roughly 5 and 10 participants. The project size distribution is highly skewed, with over 95% of the projects in FP5 having at most 15 participants while the largest project has over 100 participants. Average project size increases across the FPs, which is consistent with recommendations from evaluation studies and the stated attempts of the EU commission to reduce its administrative burden. The overall shape of the distributions, however, is remarkably similar. This may suggest that possible changes in project formation rules – including both formal policies and informal practices – did not affect the aggregate structure of the resulting research networks.

5.2. Edge properties

Relationships between the vertices of a graph are represented by the edges. So far, we have capitalised on this by using the edges to investigate properties of the vertices. As well, it is worth exploring the edges themselves more directly. An interesting and fundamental property of the edges is their mixing pattern (Newman 2003b, 2002), whether they tend to connect similar or dissimilar vertices. Here, we focus on mixing by the degrees of the vertices linked by the edge.

Mixing patterns, in terms of correlation coefficients between the degrees of linked vertices, for the FP networks are shown in Table 3. For the P-graphs, the correlation coefficients are all positive, indicating assortative mixing, with magnitudes similar to those observed in other social networks (Newman 2002). For the O-graphs, the correlation coefficients, while all positive, are miniscule, indicating that the organisations are virtually uncorrelated by degree. Quite surprisingly, the correlation coefficients in the bipartite graph show a third mixing pattern, with all coefficients small in magnitude, but negative. Based on the standard errors, the correlation coefficients for the bipartite- and P-graphs significantly differ from zero in a statistical sense, but it is doubtful that the difference is significant in a practical sense.

Table 3: Mixing patterns, FP1–FP5

Framework Programme	bipartite graph	O-graph	P-graph
FP1	-0.133 (0.003)	0.019 (0.007)	0.319 (0.003)
FP2	-0.038 (0.004)	0.016 (0.002)	0.151 (0.003)
FP3	-0.019 (0.002)	0.056 (0.002)	0.184 (0.002)
FP4	-0.030 (0.002)	0.050 (0.001)	0.194 (0.001)
FP5	-0.020 (0.002)	0.032 (0.001)	0.167 (0.001)

Note: Standard error is shown parenthetically. Values calculated using bootstrapping with 100 samples.

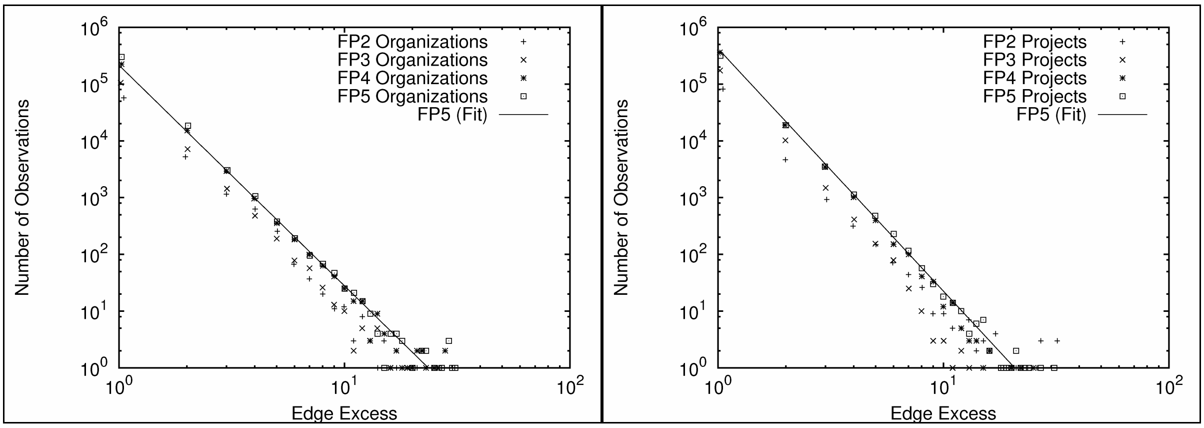
Some interpretation of the lack of strong correlation in the O-graph can be given under the assumption that the vertex size of an organisation correlates positively with its degree in the network. In this case, assortative mixing would indicate that organisations of similar size tend to take part in the same projects, while disassortative mixing would indicate that organisations of dissimilar size tend to take part in the same projects. Both types of interaction are known to occur, although the homogeneity of the network with respect to mixing patterns is less clear: there may be no tendency for degree correlation throughout the network, or the mixing patterns may vary in different subnetworks. Future work on subnetworks will shed light on this point.

The assortative mixing observed in the P-graphs indicates that projects tend to be connected to other projects of similar size. Since a connection between two projects exists when an organisation takes part in both projects, this further indicates that organisations tend to take part in projects of a particular scale with some consistency. A possible interpretation of assortative mixing amongst projects is thus that projects with similar goals tend to be organised similarly; as a corollary, particular types of projects may have specific requirements reflecting the needs of the field, such as organisation competencies or resource requirements in order for project members to be effective.

Thus far, we have focused simply on the presence or absence of links between network nodes. For a link between two organisations or projects it is sufficient to have just one project or organisation, respectively, in common, but of course there could be additional commonality. Define the *edge excess* as the number of projects or organisations beyond the minimum of one needed to form the edge.

In Figure 5, the edge-excess distribution is shown for P- and O-graphs of FP2-5. There is an almost perfect power-law behaviour in all cases, with maximum edge excesses of approximately 30.

Figure 5: Edge-excess distribution of FP2-FP5 organisation and project projection



Note: Exponent of the fit line in the organisation projection is $\alpha = 3.9 \pm 0.1$. Exponent of the fit line in the project projection is $\alpha = 4.3 \pm 0.1$. Standard error is calculated using bootstrapping with 1000 samples. Fit lines for earlier FPs are similar (not shown).

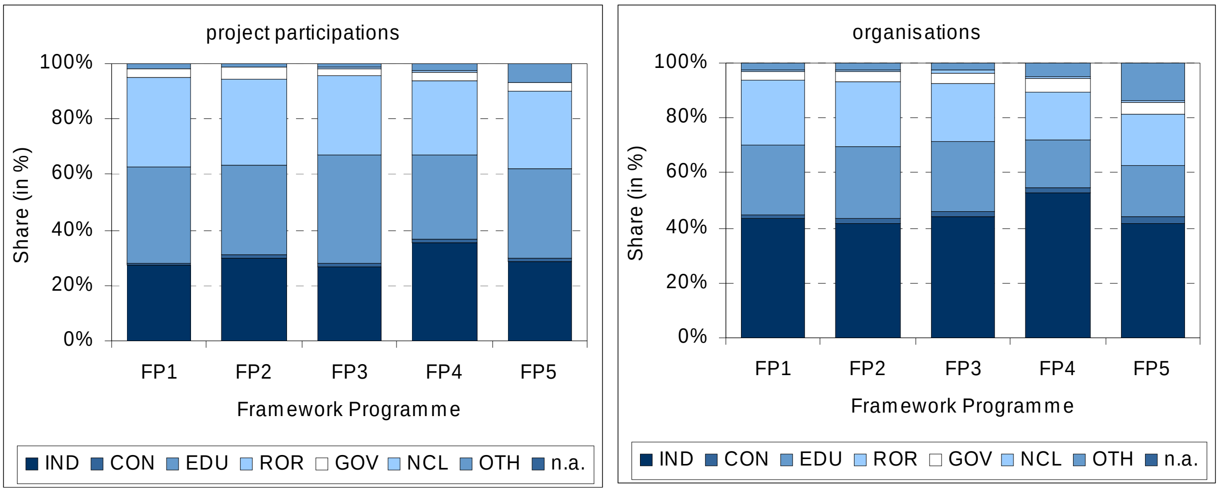
The presence of exceptionally high excesses in the P-graphs may be caused by memory effects due to prior collaborative experience. This can be validated by examining the dynamical properties of the networks. Also, a greater edge excess may result from the fact that organisations are active in a wider set of complementary activities. In this case, intra-organisational links and knowledge flows may also be of importance, as search for potential partners may be influenced by the collaboration behaviour of other actors within an organisation. The fact that a sizeable number of organisations collaborate more than once in each FP indicates that there is a kind of robust backbone structure in place, which may constitute the core of the European Research Area.

5.3. Vertex properties

5.3.1. O-graphs

The networks at hand include a diverse set of actors. We start with analysing actors' identities within FPs. Figure 6 displays the distribution of organisation types for each of the five FPs. The figure on the left is generated from the total set of project participations, while the figure on the right is based on counts of distinct organisations. Both figures show that the vast majority of participants in EU projects are firms, universities or research organisations. Measured by project participations (left-hand side), shares are quite stable over time, but this picture changes markedly when we count by distinct organisations (right-hand side).

Figure 6: Distribution of organisation types in the organisation networks, FP1–FP5



Note: IND … industry, CON …consulting, EDU … universities and other higher education, ROR … non-university research, GOV … governmental, NCL … non-commercial (associations, NGOs, etc.), OTH … other, n.a. ... not available.

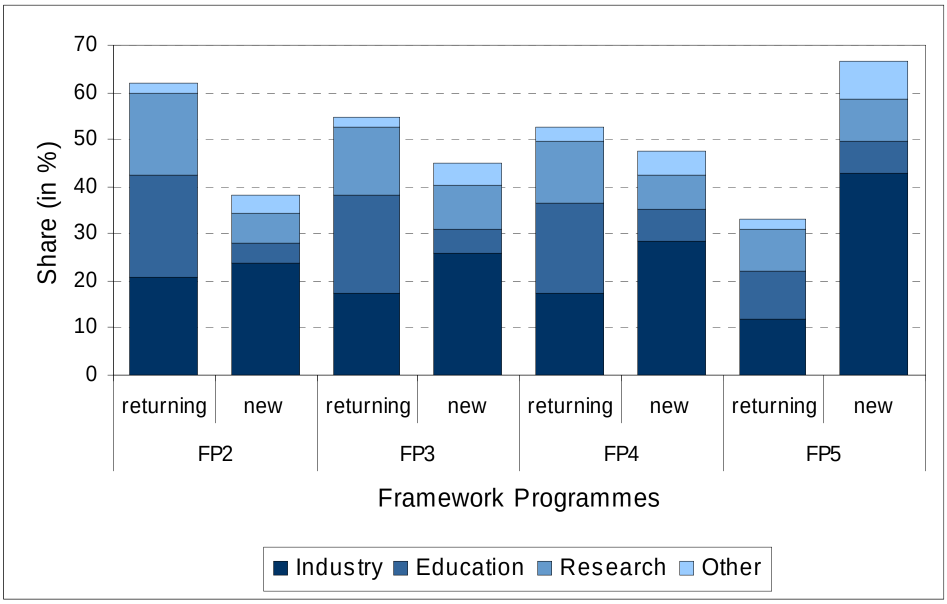
This implies a considerably higher participation intensity by actors from research than from industry. Averaging across all FPs, an industrial actor participates in a mean (standard deviation) number of 1.9 (3.4) projects, a research organisation in 4.0 (8.8) projects and a university in 4.6 (7.3). We interpret this result as evidence for greater fluctuation among industry participants and different organisational attitudes towards the kind of research captured by our data. The enormous variation of our results indicates considerable heterogeneity within the different groups of actors.

Universities and research organisations mainly conduct exploratory (re)search in the 'open science' mode, while firms focus on exploitation governed by the norms of 'proprietary technology' (see Dasgupta and David 1994). The disclosure rules stipulated in the EU FPs work well with exploratory research in the open science mode, but are ill-suited for exploitation. Exploiting existing capabilities that are critical for industrial competitiveness requires secrecy and is therefore typically funded internally. Indeed, research on Finnish firms has shown that these often set up parallel, internal projects in which they exploit results obtained in EU-funded research (Luukonen 2002).

While actors in science mainly conduct exploratory research, it constitutes a smaller part of firms' R&D activities (albeit one that is critical for long-term competitiveness). Accordingly, EU projects are a natural way of funding academic research which is reflected in greater participation intensity by scientific actors. In contrast, firms participate in fewer projects, mainly to acquire knowledge critical for longer-term success or to create future markets, e.g. by setting standards in network technologies.

Next, we explore the identity of actors between FPs. Organisations may not only participate in multiple projects within FPs, but also across them. At the same time, new actors may enter the scene. Figure 7 displays the respective relative shares and breaks them down by organisation type. It reveals two overarching trends:

Figure 7: Returning and new actors, breakdown by organisation type, FP1–FP5



First, the share of actors participating in the preceding FP is larger than the share of new actors until FP4. This is even more remarkable as the number of organisations per FP is rising steeply (see Table 2a). FP5 breaks with this trend in that more than two-thirds of the participants are new. A partial explanation is offered by the increasing integration of the candidate countries that joined the EU in its latest accession round. It will be interesting to explore other driving forces in future work.

Second, in all cases, the majority of new actors are firms. In contrast, the majority of universities and research organisations overlap between FPs, indicating considerably greater stability among these actor groups.

Figure 7 thus shows that the growing size and widening scope of the FPs has attracted a growing share of new actors, most of which are firms. At the same time, there is sizeable stability among a subset of participants. About one third of the actors taking part in each of the preceding FPs also participate in FP5. This suggests that there has been a growing organisational and social infrastructure in science and technology over the past two decades that may form the core of the present day ERA. More detailed dynamic analyses may provide interesting insights.

We next focus on the main players in this potential core of ERA. Who are they? Do they reflect the participation patterns identified in Figure 6? To shed light on these questions, we need to operationalise importance or prominence. One way to think of importance is to look at how much research an organisation carries out, approximated by how active it is in the FPs. Presumably, organisations that are able to participate in many projects carry out a greater share of the total research than less active organisations. They create more knowledge and artefacts, and are more visible. We can thus create a ranking of organisations based on the number of projects they participate in.

Table 4 shows the resulting ten most active organisations per FP. In each of the FPs, the most active organisations are predominantly large research centres, in particular the various subunits of the French CNRS. Other research organisations that rank among the Top 10 in more than one FP include the German Aerospace Center, the division 'Earth and Environment' of the Italian CNR, the French Commissariat à l'Energie Atomique (CEA), in particular its material science division, and the Dutch TNO as well as the Harwell Laboratory of the UK Atomic Energy Agency in the early FPs.

A number of industrial firms also rank among the Top 10. These include Fiat, Siemens and DaimlerChrysler in the recent FPs, and BEA Systems and Bull Europe in the early ones. The rankings are completed by the natural science and engineering faculties of some of the best known universities in Europe. These include Imperial College London, Oxford University,

Table 4: Number (#) of projects per organisation, Top 10, FP1–FP5

Rank	FP5 Organisation	#	Org-type	FP4 Organisation	#	Org-type	FP3 Organisation	#	Org-type	FP2 Organisation	#	Org-type	FP1 Organisation	#	Org-type
1	CNRS/Sciences de la vie	223	ROR	CNRS/Sciences de la vie	196	ROR	CNRS/Sciences de l'univers	170	ROR	CNRS/Sciences de la vie	94	ROR	Commissariat à l'Energie Atomique	86	ROR
2	CNRS/Sciences de l'univers	164	ROR	CNRS/Sciences de l'univers	187	ROR	CNRS/Sciences de la vie	153	ROR	CNRS/Sciences de l'univers	72	ROR	UKAEA/Harwell Laboratory	68	ROR
3	Fiat/Centro Richerche	153	IND	Fiat/Centro Richerche	112	IND	CNRS/Sciences physiques et mathematiques	135	ROR	TNO	70	ROR	TNO	65	ROR
4	HHG/German Aerospace Center	118	ROR	CNR/Earth and Environment	107	ROR	CNRS/Sciences chimiques	112	ROR	CNRS/STIC	68	ROR	Universite Catholique de Louvain	64	EDU
5	CNRS/Information et communication	113	ROR	HHG/German Aerospace Center	107	ROR	ImperialCL/Faculty of Engineering	84	EDU	CNRS/Sciences physiques et mathematiques	65	ROR	CNRS/Sciences de l'univers	64	ROR
6	CEA/Direction des Sciences de la Matiere	111	ROR	CSIC/Biology and Biomedicine	100	ROR	UP XI/Faculte des Sciences	84	EDU	Siemens AG	60	IND	Imperial College London	61	EDU
7	CNR/Earth and Environment	106	ROR	JRC/Institute for Environment and Sustainability	93	ROR	CEA/Direction des Sciences de la Matiere	82	ROR	HHG/Research Center for Environment and Health	59	ROR	BAE Systems Electronics Ltd	45	IND
8	Siemens AG	85	IND	Siemens AG	86	IND	CU/School of Physical Sciences	76	EDU	Bull Europe	58	IND	Risø National Laboratory	42	ROR
9	Uni Stuttgart/Faculty of Engineering	83	EDU	ImperialCL/Faculty of Engineering	84	EDU	CNRS/Information et communication	75	ROR	CNR/Earth and Environment	57	ROR	INRA/Centre de Recherche de Paris	39	ROR
10	DaimlerChrysler AG	82	IND	SotonU/Engineering, Science and Mathematics	79	EDU	OU/Mathematical and Physical Sciences Division	72	EDU	UKAEA/Harwell Laboratory	56	ROR	Centre d'Etudes de l'Energie Nucleaire	39	ROR
10							SotonU/Engineering, Science and Mathematics	72	EDU	BAE Systems PLC	56	IND			

Note: CEA ... Commissariat à l'Energie Atomique, CNR ... Consiglio Nazionale delle Ricerche, CNRS ... Centre National de la Recherche Scientifique, CNRS/STIC ... CNRS/Sciences et technologies de l'information et de la communication, CSIC ... Consejo Superior de Investigaciones Científicas, CU ... Cambridge University, HHG ... Helmholtz-Gemeinschaft, ImperialCL ... Imperial College London, INRA ... Institut National de la Recherche Agronomique, JRC ... Joint Research Centre, OU ... Oxford University, SotonU ... Southampton University, UP XI ... Université Paris-Sud XI, UKAEA ... United Kingdom Atomic Energy Authority.

Table 5: Central organisations, Top 10, FP1–FP5

Rank	FP5 Organisation	FP5 Org-type	FP5 Score	FP4 Organisation	FP4 Org-type	FP4 Score	FP3 Organisation	FP3 Org-type	FP3 Score	FP2 Organisation	FP2 Org-type	FP2 Score	FP1 Organisation	FP1 Org-type	FP1 Score
1	CNRS/Sciences de la vie	ROR	13	SotonU/Engineering, Science and Mathematics	EDU	11	CNRS/Sciences de la vie	ROR	9	CNRS/Sciences chimiques	ROR	18	TNO	ROR	16
2	Fiat/Centro Richerche	IND	16	HHG/German Aerospace Center	ROR	25	SotonU/Engineering, Science and Mathematics	EDU	17	Siemens AG	IND	26	Imperial College London	EDU	36
3	HHG/German Aerospace Center	ROR	18	CNRS/Sciences de la vie	ROR	36	CNRS/Sciences de l'univers	ROR	25	HHG/Research Center for Environment and Health	ROR	52	Universite Catholique de Louvain	EDU	39
4	CNRS/Information et communication	ROR	23	CNRS/Sciences de l'univers	ROR	43	UP XI/Faculte des Sciences	EDU	28	TNO	ROR	61	Risø National Laboratory	ROR	57
5	CNR/Earth and Environment	ROR	31	JRC/Institute for Environment and Sustainability	ROR	48	CNRS/Sciences chimiques	ROR	34	Bull Europe	IND	62	Technical University of Denmark	EDU	71
6	AUTH/Faculty of Technology	EDU	31	CNR/Earth and Environment	ROR	49	CEA/Direction des Sciences de la Matiere	ROR	35	Energy Research Centre of the Netherlands	ROR	78	Universiteit Twente	EDU	93
7	CEA/Direction des Sciences de la Matiere	ROR	43	LU/Institute of Technology	EDU	68	OU/Mathematical and Physical Sciences Division	EDU	53	BAE Systems PLC	IND	81	University of Liege	EDU	98
8	CNRS/Sciences de l'univers	ROR	45	CEA/Direction des Sciences de la Matiere	ROR	70	CNRS/Sciences physiques et mathematiques	ROR	56	CNR/Earth and Environment	ROR	96	Forschungszentrum Jülich	ROR	103
9	ImperialCL/Faculty of Engineering	EDU	49	Fiat/Centro Richerche	IND	73	ImperialCL/Faculty of Engineering	EDU	56	INESC ID Lisboa	ROR	97	Siemens Nixdorf Informationssysteme AG	IND	107
10	JRC/Institute for Environment and Sustainability	ROR	51	CNRS/Sciences pour l'ingenieur	ROR	102	CNRS/STIC	ROR	61	Trinity College Dublin	EDU	101	Energy Research Centre of the Netherlands	ROR	108

Note: AUTH ... Aristoteles University of Thessaloniki, CEA ... Commissariat à l'Energie Atomique, CNR ... Consiglio Nazionale delle Ricerche, CNRS ... Centre National de la Recherche Scientifique, CNRS/STIC ... CNRS/Sciences et technologies de l'information et de la communication, LU ... Lund University, INESC ... Instituto de Engenharia de Sistemas e Computadores, HHG ... Helmholtz-Gemeinschaft, ImperialCL ... Imperial College London, INRA ... Institut National de la Recherche Agronomique, JRC ... Joint Research Centre, OU ... Oxford University, SotonU ... Southampton University, UP XI ... Université Paris-Sud XI.

Cambridge University, the University of Paris XI, and the Catholic University of Louvain, as well as, perhaps unexpectedly, Southampton University and the University of Stuttgart.

The ranking of the most active organisations per FP reflects the changing funding priorities (see Figure 1). In FP1 and FP2, the most active organisations predominately conduct research on (nuclear) energy and IT. In the more recent FPs, the most active organisations mainly focus on life sciences, engineering and materials sciences (in particular automobiles and aerospace), and environmental sciences. Interestingly, organisations focusing on ICT do not dominate the rankings, even though the majority of research funding goes to this thematic area.

Defining importance by activity, however, does not capture the relational information embedded in network structures. From a network perspective, an organisation is important if it occupies a central position. Participating in many projects is not a sufficient condition; centrality depends on to whom actors are connected. If we thus want to exploit the added value of our relational information, centrality is the theoretically appropriate measure to identify main actors. Unfortunately, theory offers limited guidance on the most appropriate measure of centrality, so we proceed pragmatically.

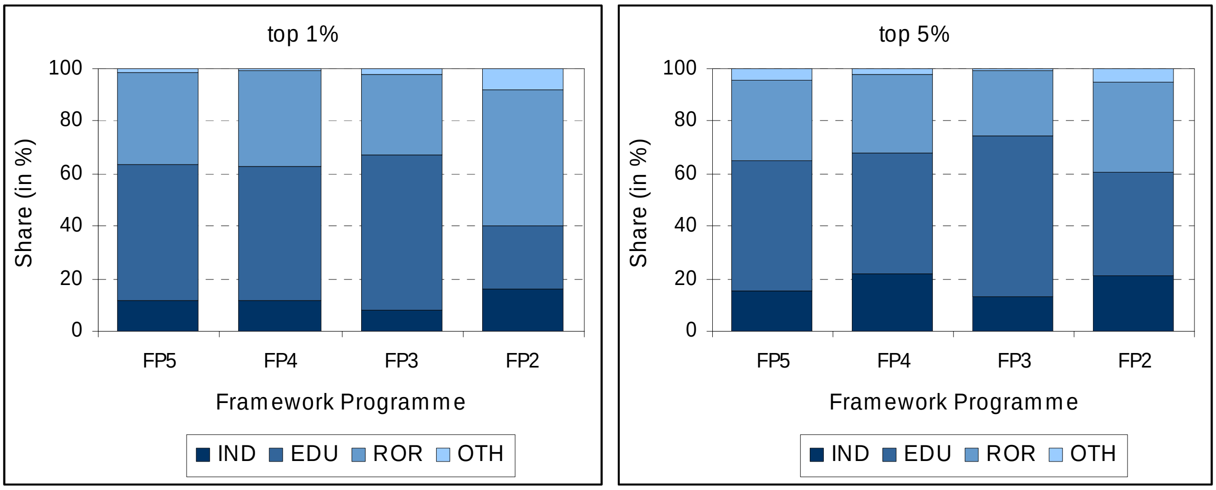
Using each of the four centrality measures described in Section 3.3.4., we compute centrality weights for each vertex. This is done with Pajek (Batagelj and Mrvar 2003), a free application for the descriptive analysis of large networks. We then compute an aggregate score of vertices by taking the unweighted sum of their individual rankings according to each centrality index. The intuition for our methodology is that truly central vertices should rank prominently along each of the dimensions quantified by the centrality indices.

Table 5 displays the ten most central organisations in each FP, identified in this manner. Median aggregate centrality is lowest in FP5, followed by FP3. As expected, the centrality rankings do not coincide with the activity rankings.

Among the Top 10, universities rank higher, while firms move down the ranking. For instance, Southampton University is most central in FP4 and second most central in FP3, up from tenth by activity in both. Lund University is 24th by activity in FP4 and the University of Twente ranks only 55th in FP1. In contrast, Siemens only ranks 85th by aggregate centrality in FP4 (19th in FP5) and BAE Systems is 30th in FP1. There is similar variation in the ranking of research organisations, but no discernible trend.

To see whether these observations reflect a general trend in the data, we need to compare larger subsets. Since we are interested in the potential core of ERA, we identify the organisations ranked in the top one and the top five percentile by participation and break them down by the three main organisation types. Figure 8 displays the resulting distributions.

Figure 8: Most active organisations by organisation type, FP2-FP5



Note: IND … industry, EDU … universities and other higher education, ROR … non-university research, OTH … consulting, governmental, non-commercial and other.

The figure shows that the extreme tail of the activity distribution of participants in FP3 to FP5 comprises predominantly universities (45-60%) and research organisations (25-35%), while industry accounts only for a minor share (10-20%). This is particularly pronounced among the most active 1%. Among the top five percentile, there are relatively more firms, fewer universities (in FP4 and FP5) and fewer research organisations (FP2 to FP5). FP2 is special as research organisations are particularly prominent in the top one percentile.

These results show that the potential core of ERA, as identified by participation intensity in the EU FPs, is heavily skewed towards actors from science and, above all, universities. This is in stark contrast with the general participation patterns identified in Figure 6. While accounting for much total research activity, relatively few firms are part of the core.

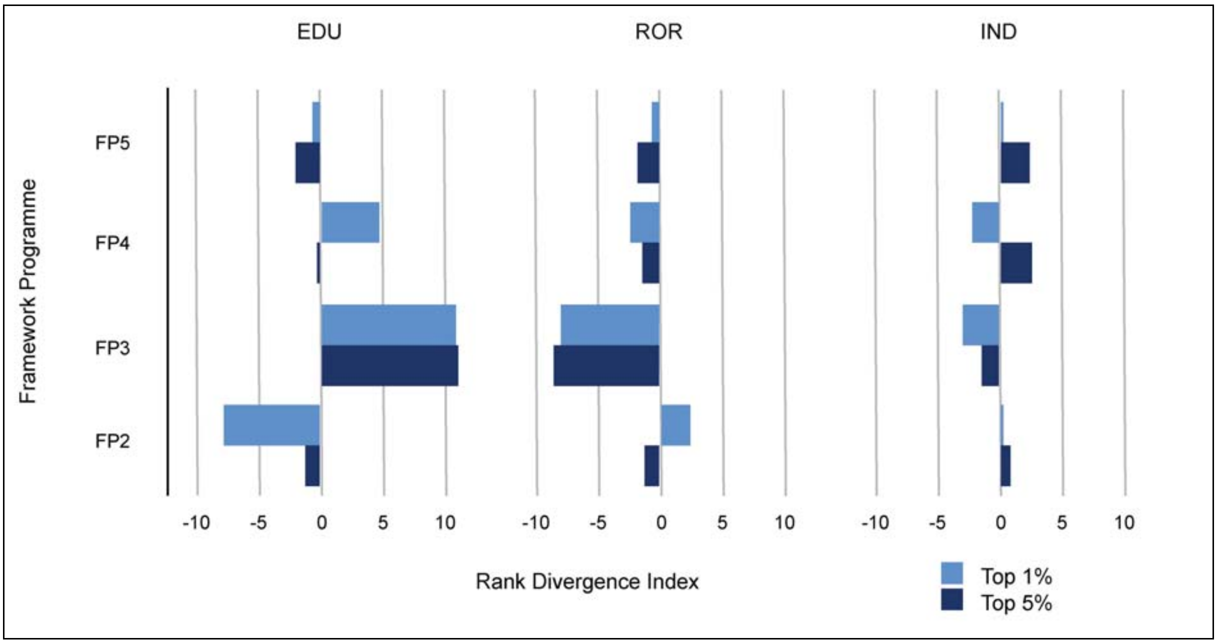
Next, we check whether this conclusion needs to be modified if we consider the relational information provided by the centrality rankings. As universities and research centres participate in more projects on average than firms, they are linked to more actors and we hence expect them to be relatively more important than participation alone would indicate. This expectation is partially borne out by the top 10 ranking reported above.

To test this, we compare the respective centrality and participation rankings for each organisation type. Tie values lead to different numbers of organisations in the resulting sets, so it is necessary to rescale proportionately, allowing comparison within the FPs by the difference of the values. Further, we would like to compare across the FPs, so we need to rescale the differences as well. By scaling to a size of 100, we obtain a rank divergence index that is simply the difference between the percentages of the relevant organisations in the respective percentile of either ranking. This index is positive for organisation types that are more highly ranked by centrality and negative for those that are more highly ranked by participation.

Figure 9 exhibits the results of this procedure. It displays index values for the three main sets of actors for FP2 to FP5. To facilitate comparison across FPs, index values are presented in three columns, each representing one organisation type. To identify variation in the extreme tail of the ranking distributions, we plot separate values for the top one and the top five percentiles.

Absolute index values show that the two ranking methods yield greater differences for actors from science than for industrial actors. This suggests that network structure has a greater impact on how scientific actors are positioned in the centrality rankings. This observation is consistent with the result that actors from science participate in more projects than firms and hence may have more complex relations in the networks we study.

Figure 9: Rank divergence index by organisation type, FP2-FP5



However, Figure 9 refutes our conjecture on the direction of the differences between the ranking methods. Across FP2-FP5, in the potential core of ERA there are fewer actors from science and more actors from industry by centrality than by participation ranking. With the exception of the top one percentile in FP2, index values for research organisations are negative.

Universities' index values are negative in FP5 and FP2, positive in FP4 and robustly positive in FP3. Index values for industry are positive in FP5 and FP2, and negative in FP3. In FP4, the rankings based on the top one and the top five percentile yield opposite results.

These results reveal considerable variation in the way the different sets of actors are embedded in the respective R&D networks in FP2 to FP5. Differences in index values for the top one and the top five percentile rankings are due to variations in the distribution of organisations in the respective percentiles. This is particularly obvious if the index values produce opposite results, as in the case of industry in FP4 and research organisations in FP2. For example, in FP4 there are relatively few firms in the top one percentile of the centrality ranking, yielding a negative index value. In contrast, they are substantially overrepresented in the top five percentile, yielding a positive index value. Thus, there is clearly a non-linear relationship between the two ranking methods caused by structural properties of the networks that we do not understand yet.

5.3.2. P-graphs

As in the O-graphs, we can discriminate sets of vertices in the P-graphs. The EU uses different instruments to fund research in the FPs, which are likely to have different structural properties. Table 6 shows the average project size of the contract types we included in our networks. Since shared-cost projects were the only instrument in FP1 and FP2, we do not list them separately (for details, see Table 2b). To avoid biased estimates, we only consider multiple partner projects. Note that due to the scale-free distribution of the data (see Figure 4), some care is warranted in interpreting the results.

Table 6: Average project size by contract type, FP3-FP5

		CSC	CRS	CON	DEM	THN
FP5	# multiple partner	5255	658	141	170	513
	# available total	5271	659	173	172	534
	mean (StDev)	8.2 (4.2)	8.2 (2.7)	13.0 (9.4)	7.4 (4.8)	16.6 (14.0)
FP4	# multiple partner	6956	713	242	56	125
	# available total	7156	735	574	550	128
	mean (StDev)	6.7 (3.8)	8.7 (2.8)	10.7 (7.0)	5.2 (2.0)	21.6 (14.4)
FP3	# multiple partner	3531	–	138	7	–
	# available total	4089	–	155	7	–
	mean (StDev)	6.1 (4.0)	–	12.1 (7.8)	13.7 (6.4)	–

Note: Available total number of projects per contract type listed parenthetically. Computations based on multiple partner projects. CSC ... shared-cost project, CRS ... co-operative research project, CON ... co-ordination of research actions, DEM ... demonstration project, THN ... thematic network. FP1 and FP2 only comprise CSC projects; respective values are listed in Table 2.

The table shows that the large majority of vertices in the P-graphs are shared-cost projects. A sizeable number of vertices in FP4 and FP5 are co-operative research projects aimed at SMEs; FP5 also includes more than 500 thematic networks.

Comparing projects within FPs, thematic networks and co-ordinated research actions tend to be larger on average than the remaining project types (subject to considerable variation). Comparing projects over time, the detailed breakdown shows that not all project types have grown in size as identified as the average trend in Table 2b. Co-operative research projects and co-ordinated research actions have stable average sizes, and thematic networks may have become smaller in FP5.

A fundamental question that has received particular prominence with the introduction of new instruments in FP6 is which projects are pivotal for the coherence of R&D networks in Europe. To shed some light on this point, we identify central projects by applying the same methodology as with the O-graphs. We consider the four centrality measures and compute aggregate rankings. Again, the intuition is that central projects should be well positioned according to all dimensions of centrality captured by the indices. Table 7 shows the ten most central projects for FP3-FP5. Qualitatively, FP1 and FP2 produce similar results, but are not listed for lack of space. Aggregate centrality scores are considerably higher than in the O-graphs (see Table 5), indicating greater heterogeneity in individual centrality rankings.

Projects ranked among the Top 10 tend to be considerably larger than the average project in each FP. This is an expected result, in particular as we assume projects to be fully connected subgraphs. However, linkage structure matters, as most of the largest projects are not among the most central ones (only in FP3, the largest project is among the most central, ranked first). In terms of project type, only two thematic networks and two concerted research actions are among the most central projects, with the remainder being shared-cost projects. This suggests that the different instruments may address different groups of actors.

These observations hold if we consider a greater part of the extreme tail of the centrality ranking distribution. Repeating the same analyses as with the O-graphs, we obtain the following results.

First, association between project centrality and project size rankings is unexpectedly weak. The rank-based Spearman-Rho correlation coefficient produces values ranging from 0.3-0.4, indicating a medium-strong correlation. Thus, large projects are not necessarily central projects, implying that large projects include a fair number of peripheral actors. This may have a positive effect on knowledge dissemination, as large projects may succeed in connecting cen-

Table 7: Central projects, Top 10, FP3-FP5

Rank	FP5					FP4					FP3				
	PrAcr	Title	#	Ctype	Score	PrAcr	Title	#	Ctype	Score	PrAcr	Title	#	Ctype	Score
1	EESD	Mediterranean ocean Forecasting System: Toward Environmental Predictions	48	CSC	22	MAST 3	Mediterranean Targeted Project II-Mass Transfer and Ecosystem Response	56	CSC	7	BIOTECH 1	Plant molecular genetics for an environmentally compatible agriculture	78	CSC	4
2	GROWTH	The European network for superconductivity, scenet-2	74	THN	26	MAST 3	Canary Islands Azores Gibraltar Observation	45	CSC	49	HCM	AB initio (from electronic structure) calculation of complex processes in materials	39	CSC	20
3	LIFE QUALITY	Concerted action on mitochondrial biogenesis and disease	41	CON	68	BIOTECH 2	Extremophiles as cell factories	40	CSC	62	HCM	Molecular dynamics and monte carlo simulations of quantum and classical systems	33	CSC	29
4	EESD	European Network for Biodiversity Information	66	THN	152	MAST 3	Ocean Margin EXchange II - Phase I	32	CSC	84	HCM	Non linear phenomena in microphysics of collisionless plasmas. Application to space and laboratory plasmas.	18	CSC	31
5	EESD	Flume Facility Co-operation Network for Biological Benthic Boundary Layer Research	21	CON	153	BIOTECH 2	Characterising and engineering abscisic acid action	12	CSC	124	MAST 2	Ocean margin exchange	33	CSC	31
6	EESD	Atmospheric Deposition and Impact of pollutants, key elements and nutrients on the Open Mediterranean Sea	28	CSC	153	MAST 3	Ocean Margin Exchange II - Phase II	27	CSC	127	AIR	A Multidisciplinary Research Network Study and Improve the Abiotic Stress Tolerance of European Agricultural Crops	44	CON	83
7	EESD	The Airborne Platform for Earth observation Infrastructure	20	CSC	158	MAST 3	Baltic Sea System Study	48	CSC	144	HCM	European Network on AntiBody Catalysis	24	CSC	92
8	EESD	Development of an Information Technology Tool for the Management of European Southern Lagoons under the influence of river-basin runoff	22	CSC	160	ENV 2C	European forum on integrated environmental assessment	26	CSC	160	HCM	Structure and reactivity of molecular ions	27	CSC	94
9	EESD	European catchments, catchments changes and their impact on the coast	25	CSC	160	MAST 3	Azores mid-oceanic ridge ecosystem studies: an integrated research programme on deep sea hydrothermal transfers and fluxes	19	CSC	166	HCM	Metal ion-nucleic acid interactions and antitumour drugs	23	CSC	106
10	LIFE QUALITY	Gene flow from transgenic plants: evaluation and biotechnology	12	CSC	167	BIOTECH 2	Molecular characterization of cold-active enzymes from psychrophilic microorganisms as the basis for novel biotechnology	11	CSC	180	HCM	Selective processes and catalysis involving small molecules	19	CSC	132

Note: PrAcr ... specific programme acronym, # ... partners/project, ctype ... contract type. EESD ... Energy, Environment and Sustainable Development, GROWTH ... Competitive and Sustainable Growth, LIFE QUALITY ... Quality of Life and Management of Living Resources; MAST ... Marine Sciences and Technologies, BIOTECH ... Biotechnology, ENV ... Environment and Climate; HCM ... Human Capital and Mobility, AIR ... Agriculture and Agro-Industry including Fisheries.

tral and peripheral actors. At the same time, it suggests that to create pivotal nodes, it is at least as important to assemble well-connected actors as it is to assemble many actors. It will be interesting to explore these highly policy-relevant questions in future work.

Second, the association pattern between centrality and size ranking is also reflected in the distribution of contract types in the top one and top five percentiles of the respective ranking distributions. Being larger on average, thematic networks and co-ordinations of research actions are heavily overrepresented in the extreme tail of the project size rankings, compared to their overall share in the FPs. In contrast, they feature markedly less prominently in the top one and top five percentiles of the project centrality ranking distributions. In the FP4 and FP5 P-graphs, co-ordinated research actions have rank divergence index values of about -10 and thematic networks of -20 to -40, while shared-cost projects range from +10 to +40. These differences are considerably larger than in the O-graphs and underline that different project types appear to be embedded differently in the respective P-graphs, attracting different groups of actors.

Finally, the most central projects only partially reflect the funding priorities of the FPs. This is evident already from the Top 10 ranking shown in Table 7, in which biotechnology and life science projects are positioned centrally, while information technology, engineering and materials science projects are virtually absent. Instead, a surprisingly large number of projects are in environmental (marine) sciences and technologies.

These observations can be substantiated for FP4 and FP5 by assigning the projects in the extreme tail of the centrality ranking distributions to the main thematic priorities (see Figure 1) on the basis of their subject indices. These are standardised keywords that are available for all projects in the sysres EUPRO database. Roughly 40% of the projects are in the priority environment, 25% are in industrial manufacturing, 20% (FP4) to 25% (FP5) are in life sciences, whereas less than 10% are in ICT and energy, respectively. Within these priorities, marine sciences and technologies and aeronautics are particularly prominent. Especially the former are usually not considered as core areas of EU funded research. This result is a strong indication that EU funded research in different thematic areas is organised differently, quite possibly reflecting the technological and economic needs of the field.

Our results on central projects only partially coincide with our results on central actors. This further underlines the need to gain a much better understanding of the internal structure for projects based on which actors participate in what kind of projects and to identify coherent substructures in the global networks.

6. Conclusions and directions for future research

In this work, we have investigated the structure of R&D collaboration networks determined from research projects funded in the EU Framework Programmes. The networks are substantial in terms of size, complexity, and economic impact. We observe numerous characteristics known from other complex networks, including scale-free degree distributions, small diameters, and high clustering. The networks thus exhibit the small-world property, which has been identified in theoretical work as conducive to collective knowledge creation and knowledge transmission in exploration networks. Other features in common throughout the FPs include the typical project sizes and the overall shapes of the various distributions observed. Presumably, network formation mechanisms are similar for all FPs despite changes in governance rules.

Two findings suggest the presence of a stable core of actors in science and technology since the early FPs: there is a significant overlap in participants for consecutive FPs and there is recurring collaboration amongst the same organisations within FPs. This core may constitute the backbone of the present day European Research Area. Moreover, the increasing clustering coefficient suggests that integration between collaborating organisations has increased over time, indicating that Europe has already been moving towards a more closely integrated European Research Area in the earlier Framework Programmes.

We observe assortative mixing patterns in the project networks with positive correlation by degree. As organisations are the medium through which project degree correlations occur, this indicates that organisations tend to participate in similarly sized projects. Assuming that organisations operate in coherent fields of activity, this may reflect the needs of their field. Unexpectedly, we find no degree correlation in the organisation networks. The reason for this is not clear. It may be that there simply is no trend or that subnetworks follow very different mixing patterns.

Further results stem from investigating vertex properties. We find that the majority of participants are firms, but that universities and research organisations display greater participation intensity and are positioned more prominently in the networks analysed. We observe and characterise considerable variation in how the main sets of actors, universities, research organisations and firms are embedded in networks studied, both within and between Framework Programmes.

Including the relational information contained in the networks shows that vertex size does not imply centrality. This suggests that a policy of creating larger projects may not be fully appropriate to foster networking and the connectivity of R&D collaboration networks in Europe. Rather, projects need to include pivotal actors, which seems to have been the case only partially in first five Framework Programmes.

We also find that the most central projects do not reflect the main funding priorities of the FPs. We interpret this as an indication of differences in how research in specific thematic areas is organised, possibly due to field-specific technological and economic needs.

The present work points to considerable future work, of both empirical and conceptual nature. At the empirical level, it is clear that we need to refine our understanding of the substructure of the networks, in particular how organisations interact within projects. This includes identifying thematically homogeneous subnetworks and subgroups that are homogeneous in terms of structural properties and organisational mixing patterns. Moreover, the additional information included in edge weights needs to be integrated into structural investigations. Another major route of inquiry is the dynamic analysis of network formation and network configuration. This should also yield information on how the networks have been shaped by external constraints, in particular the governance rules.

Perhaps more fundamentally, there are many open questions at the conceptual level. Networking activities are publicly funded because they are expected to fulfil specific functions, e.g. knowledge creation and knowledge diffusion. Are the network structures that emerge well-suited for these functions? Do different network functions require different network structures? Does this introduce tensions between conflicting objectives, e.g. efficiency and equity? How do structure and function interact? To what extent can complex networks that involve a strong element of self-organised behaviour by decentralised actors be influenced through external stimuli, in particular by governance rules? Isolating relationships between network structure and function and identifying the scope for directing networks towards desirable structures will provide valuable guidance for policy makers in improving existing instruments and designing future ones.

Acknowledgements

We acknowledge support from the Portuguese Fundação para a Ciência e a Tecnologia under Bolsa de Investigação SFRH/BPD/9417/2002 and Plurianual CCM. TRS gratefully acknowledges the support of CCM at Madeira Math Encounters XXX, where portions of this work were done. An earlier version of this paper was presented at the International Schumpeter Conference 2006. We would like to thank participants for helpful comments.

References

Albert, R. and A.-L. Barabasi (2002), 'Statistical mechanics of complex networks', *Reviews of Modern Physics*, **74** (1), 47-96.

Barabasi, A.-L. and R. Albert (1999), 'Emergence of scaling in random networks', *Science*, **286**, 509-12.

Barber, M., A. Krueger, T. Krueger and T. Roediger-Schluga (2006), 'The network of EU-funded collaborative R&D projects', *Physical Review E*, **73**, 036132 (arXiv:physics/0509119).

Barker, K. and H. Cameron (2004), 'European Union science and technology policy, RJV collaboration and competition policy', in Y. Caloghirou, N.S. Vonortas and S. Ioannides (eds.), *European Collaboration in Research and Development*, Cheltenham, UK and Northampton, MA, US: Edwar Elgar, 154-86.

Batagelj, V. and A. Mrvar (2003), 'Pajek – Analysis and visualization of large networks', in M. Jünger and P. Mutzel (eds.), *Graph Drawing Software*, Berlin: Springer, 77-103.

Bollobás, B. (2001), *Random Graphs,* 2nd edn, Cambridge: Cambridge University Press.

Bonacich, P. (1987), 'Power and centrality: A family of measures', *American Journal of Sociology*, **92**, 1170-82.

Bonacich, P. (1972), 'Technique for analyzing overlapping memberships', in H.L. Costner (ed.) *Sociological methodology*, San Francisco: Jossey-Bass, 176-85.

Bornholdt, S. and H.G. Schuster (eds) (2003), *Handbook of Graphs and Networks*, Weinheim: Wiley-VCH.

Breschi, S. and L. Cusmano (2004), 'Unveiling the texture of a European Research Area: Emergence of oligarchic networks under EU Framework Programmes', *International Journal of Technology Management*, **27** (8), 747-72.

Brin, S. and L. Page (1998), 'The anatomy of a large-scale hypertextual web search engine', *Computer Networks*, **33**, 107-17.

Caloghirou, Y. and N.S. Vonortas (2000), 'Science and Technology Policies Towards Research Joint Ventures', Final Report to the Commission, DG XII, TSER Programme, Contract No: CT97-1075, National Technical University of Athens, Laboratory of Industrial & Energy Economics, available from http://improving-ser.sti.jrc.it/default/show.gx?Object.object_id=TSER—000000000000097E&_app.page=show-TSR.html.

Caracostas, P. and U. Muldur (2001), 'The emergence of a new European Union research and innovation policy', in P. Larédo and P. Mustar (eds.), *Research and Innovation Policies in the New Global Economy*, Cheltenham, UK and Northampton, MA, US: Edward Elgar, 157-204.

CORDIS (2006a), 'Towards FP7', available from http://cordis.europa.eu/fp7/faq.htm).

CORDIS (2006b), 'FP6 Budget', available from http://cordis.europa.eu/fp6/budget.htm).

CORDIS (2006c), 'FP5 - EC Programme: Maximum Amounts and Breakdown (1998-2002)', available from http://cordis.europa.eu/fp5/src/budget.htm).

CORDIS search (2006), 'Projects Professional Search', available from http://cordis.europa.eu/search/index.cfm?fuseaction=proj.professionalSearch).

Cowan, R. (2006), 'Network models of innovation and knowledge diffusion', in S. Breschi and F. Malerba (eds.), *Clusters, Networks, and Innovation*, Oxford: Oxford University Press (forthcoming).

Cowan, R. and N. Jonard (2004), 'Network structure and the diffusion of knowledge', *Journal of Economic Dynamics and Control*, **28** (8), 1557-75.

Cowan, R. and N. Jonard (2003), 'The dynamics of collective invention', *Journal of Economic Behavior and Organization*, **52** (4), 513-32.

Cowan, R., N. Jonard and M. Ozman (2004), 'Knowledge dynamics in a network industry', *Technological Forecasting and Social Change*, **71** (5), 469-84.

Dasgupta, P. and P.A. David (1994), 'Toward a new economics of science', *Research Policy*, **23**, 487-521.

Dorogovtsev, S.N. and J.F.F. Mendes (2004), 'The shortest path to complex networks', in N. Johnson, J. Efstathiou and F. Reed-Tsochas (eds.), *Complex Systems and Interdisciplinary Science*, Singapore: World Scientific (arXiv: cond-mat/0404593).

Dorogovtsev, S.N. and J.F.F. Mendes (2002), 'Evolution of networks', *Advances in Physics*, **51**, 1079-187.

Edquist, C. (2005), 'Systems of innovation – perspectives and challenges', in J. Fagerberg, D.C. Mowery and R.R. Nelson (eds.), *The Oxford Handbook of Innovation*, Oxford: Oxford University Press, 181-208.

Efron, B. and R. Tibshirani (1993), *An Introduction to the Bootstrap,* New York: Chapman and Hall.

European Commission (2006), 'The Fourth Framework Programme', European Commission, available from http://ec.europa.eu/research/fp4.html, last update May 22, 2001).

European Commission (2005), 'Nomenclature of territorial units for statistics - NUTS Statistical Regions of Europe ', available from http://europa.eu.int/comm/eurostat/ramon/nuts/home_regions_en.html).

European Commission (2002), 'What European added value?' European Commission, available from http://ec.europa.eu/research/era/leaflet/en/era04_en.html).

European Commission (2000), 'Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions – Towards a European Research Area', COM (2000) 6, Brussels.

Goldstein, M.L., S.A. Morris and G.G. Yen. (2004), 'Problems with fitting to the power-law distribution', *European Physics Journal B*, **41** (2), 255-58.

Guzetti, L. (1995), *A Brief History of European Union Research Policy,* Luxembourg: Office for Official Publications of the European Commission.

Hagedoorn, J., A.N. Link and N.S. Vonortas (2000), 'Research partnerships', *Research Policy*, **29** (4-5), 567-86.

Hodgson, G.M. (1998), 'Competence and contract in the theory of the firm', *Journal of Economic Behavior & Organization*, **35**, 179–201.

Kleinberg, J.M. (1999), 'Authoritative sources in a hyperlinked environment', *Journal of the ACM*, **46** (5), 604-32.

Larédo, P. (1998), 'The networks promoted by the framework programme and the questions they raise about its formulation and interpretation', *Research Policy*, **27** (6), 589-89.

Lundvall, B.-Å. and S. Borrás (2005), 'Science, technology, and innovation policy', in J. Fagerberg, D.C. Mowery and R.R. Nelson (eds.), *The Oxford Handbook of Innovation*, Oxford: Oxford University Press, 599-631.

Luukonen, T. (2002), 'Technology and market orientation in company participation in the EU framework programme', *Research Policy*, **31** (3), 1459-66.

McPherson, M., L. Smith Lovin and J. Cook (2001), 'Birds of a feather: Homophily in social networks', *Annual Sociological Review*, **52**, 415-44.

Newman, M.E.J. (2003a), 'The structure and function of complex networks', *SIAM Review*, **45**, 167-256.

Newman, M.E.J. (2003b), 'Mixing patterns in networks', *Physical Review E*, **67**, 026126.

Newman, M.E.J. (2002), 'Assortative mixing in networks', *Physical Review Letters*, **89**, 208701.

Peterson, J. and M. Sharp (1998), *Technology Policy in the European Union,* Houndsmill, Basingstoke and London: Macmillan.

Powell, W.W., D.R. White, K.W. Koput and J. Owen-Smith (2005), 'Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences', *American Journal of Sociology*, **110** (4), 1132-205.

Rothaermel, F.T. and D.L. Deeds (2004), 'Exploration and exploitation alliances in biotechnology: a system of new product development', *Strategic Management Journal*, **25**, 201-21.

Soete, L. and A. Arundel (eds) (1993), *An Integrated Approach to European Innovation and Technology Diffusion Policy. A Maastricht Memorandum*, Brussels: European Commission.

Strogatz, S.H. (2001), 'Exploring complex networks', *Nature*, **410**, 268-76.

Wasserman, S. and K. Faust (1994), *Social Network Analysis – Methods and Applications,* Cambridge: Cambridge University Press.
Watts, D.J. (2004), 'The "new" science of networks', *Annual Review of Sociology*, **30**, 243-70.
Watts, D.J. and S.H. Strogatz (1998), 'Collective dynamics of 'small-world' networks', *Nature*, **393**, 440-42.

Technical appendix

Fitting power laws

To describe the various distributions observed, we fit power laws to the empirical data. This appears to be a straightforward process of constructing a histogram of the data on logarithmic axes and using linear least squares to fit a line to the tail of the distribution. However, Goldstein et al. (2004) have shown that this approach is biased and can introduce substantial errors when not used with care. In essence, for a power law the end of the tail is constructed with relatively little data in comparison to the rest of the histogram, so the uncertainties at the extreme end are large. Thus the histogram is heteroscedastic, violating one of the assumptions of OLS estimation. This can be corrected by truncating the tail; Goldstein et al. also propose a more robust method based on maximum likelihood estimation. In this work, we take the simpler approach of truncating the tail.

We determine the standard errors for parameters of the power law fits using a bootstrapping procedure (Efron and Tibshirani 1993). Bootstrapping is an approach for statistical inference based on resampling from an already drawn sample. Given a sample of size N and a calculated sample statistic of interest, a so-called bootstrap sample of size N is drawn with replacement (i.e. a particular observation may be drawn multiple times or not at all) and the statistic is recalculated from the bootstrap. This is repeated a large number of times, generating a set of bootstrap statistics. The uncertainty of the statistic of interest is then estimated using the bootstrap statistics; the standard error of the sample statistic is just the standard deviation of the set of bootstrap statistics.

Eigenvector centrality

Eigenvector centrality assigns to each vertex in a graph a centrality that depends both on the number and the quality of its connections. Denoting the centrality of vertex i by x_i we define the centrality to be proportional to the average centrality of the neighbouring vertices,

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} x_j \qquad (1),$$

where λ is a constant. In an equivalent matrix form, this becomes the eigenvalue problem:

$$\mathbf{Ax} = \lambda\mathbf{x} \tag{2}$$

In a connected network, the eigenvector corresponding to the largest eigenvalue is guaranteed to have all positive elements (by the Perron-Frobenius theorem). This eigenvector can be found trivially by iterating (2) until the solution stabilises. The algorithm, implemented as PageRank in Google, is due to Brin and Page (1998) and is essentially identical in its simplest form to the eigenvector centrality long used in social network analysis (Bonacich 1987, 1972).

As posed, PageRank is oriented towards undirected networks. A generalisation to directed networks was proposed by Kleinberg (1999). The mathematical effect is similar to using the singular value decomposition of **A** instead of solving the eigenvalue problem (2). This general method is implemented in Pajek and reduces to PageRank for undirected networks, as used in this work.

Convergence of moments of power law distributions

For a power law distribution, higher-order moments will not in general be defined. To see this, consider the *n*th moment $\langle k^n \rangle$ of a power law $p(k) = ck^{-\alpha}$:

$$\langle k^n \rangle = \int_{k_0}^{\infty} k^n p(k) dk = c \int_{k_0}^{\infty} k^{n-\alpha}\, dk = \frac{c}{n-\alpha+1} k^{n-\alpha+1} \Big|_{k_0}^{\infty} \longrightarrow \infty \text{ when } n+1 > \alpha .$$

Thus, for moments of order one or more greater than the exponent, the moment diverges, so that the mean is only defined when α is greater than two.