

The interplay of universities and industry through the FP5 network

Juan A Almendral¹, J G Oliveira², L López³, Miguel A F Sanjuán¹
and J F F Mendes²

¹ Departamento de Física, Universidad Rey Juan Carlos, Tulipán s/n,
28933 Móstoles, Madrid, Spain

² Departamento de Física, Universidade de Aveiro,
Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

³ Departamento de Ingeniería Telemática y Tecnología Electrónica,
Universidad Rey Juan Carlos, Tulipán s/n, 28933 Móstoles,
Madrid, Spain

E-mail: jfmendes@fis.ua.pt

New Journal of Physics **9** (2007) 183

Received 24 January 2007

Published 28 June 2007

Online at <http://www.njp.org/>

doi:10.1088/1367-2630/9/6/183

Abstract. To improve the quality of life in a modern society it is essential to reduce the distance between basic research and applications, whose crucial roles in shaping today's society prompt us to seek their understanding. Existing studies on this subject, however, have neglected the network character of the interaction between universities and industry. Here we use state-of-the-art network theory methods to analyse this interplay in the so-called Framework Programme (FP)—an initiative which sets out the priorities for the European Union's research and technological development. In particular we study in the 5th FP (FP5) the role played by companies and scientific institutions and how they contribute to enhance the relationship between research and industry. Our approach provides quantitative evidence that while firms are size hierarchically organized, universities and research organizations keep the network from falling into pieces, paving the way for an effective knowledge transfer.

Contents

1. Introduction	2
2. Analysis of the data	3
2.1. Degree distribution	3
2.2. Shortest paths	5
2.3. Betweenness centrality	6
2.4. Clustering coefficient	8
2.5. Degree–degree correlation.	9
3. Conclusion	13
Acknowledgments	13
Appendix. Classification of participants into companies and universities	13
References	16

We address, from a complex networks viewpoint, the relationship between companies and scientific institutions and how they contribute to reduce the distance between research and applications. Since this approach requires information about real systems, we focus on the Framework Programme (FP)—an initiative which sets out the priorities for the European Union’s research and technological development. Despite the presence of many and different participants, they can be split in basically two groups: companies and universities. The first one is made of companies or other industry related participants who expect that their investments in R + D + I are profitable. The second group can be thought as the opposite: participants involved in some type of academic research for whom results do not necessarily return income. We find that the transmission of information is more efficient between universities than among companies. Furthermore, when universities are excluded from the projects, companies tend to form clusters, making communication between them difficult (if not impossible). Likewise, if we pay attention to the evolution of the FP, we see that, although the creation of collaborations is encouraged, it is mainly between universities and this is insufficient to improve the relationship between research and industry. Finally, we find that companies and universities differ in the way they establish collaborations. Large corporations are reluctant to choose small companies as partners, whereas size is not important between universities. But if we analyse how universities and companies cooperate, the result is that large universities prefer working with large companies, while companies select their collaborators between universities regardless of their sizes. These findings have potential implications for future programmes, as well as for new policies and services aiming at research, development and innovation in general.

1. Introduction

Understanding the relationship between research and industry is essential to improve the quality of life in any modern society. Ranging from faster application of new discoveries to knowing whether or where investment should be employed, this flow of knowledge between research and industry has long been of general interest. Yet, knowledge is a very special resource whose study demands new techniques. The traditional approach to resources is based on the concept of scarcity since they are usually finite. But knowledge cannot be seen this way because it grows, and the

more it is used the more it spreads [1]. In addition, existing studies on the research and industry interplay [2]–[4] have neglected its network character. Our approach consists of analyzing this issue from a complex network viewpoint [5]. Many other systems are better understood in this manner [6]–[9]. In this approach, the interaction between research and industry is best described as a network whose vertices (or nodes) represent either companies or institutions devoted to research, and each edge (or link) represents collaboration between any two of them. Hence, we can quantitatively study how research and industry influence each other, if we have access to data describing a real system.

Here, we focus our attention on the so-called Framework Programme (FP), a mechanism aiming to improve the transference of knowledge in the EU by setting out its priorities for research and technological development. The data to generate the corresponding FP network were gathered from the CORDIS website [10] by a robot. Since, currently, the 6th programme is under execution and the 7th is being planned, we focused our study on the 5th FP (FP5)—covering the period from 1998 to 2002—in order to analyse a completely finished programme. Despite the presence of more than 25 000 participants, they can be split in two major groups: companies and universities. The first is made of over 16 700 companies and other industry related participants who expect their investments in R + D + I to be profitable. The second group can be regarded as the opposite, more than 8500 participants involved in some type of research for whom results do not necessarily return immediate income (see appendix). Exploring the relationship between these two groups not only provides a good example of the interplay between structure and information flow, but also offers a glimpse on how research links with innovation and if the distance between basic research, applications and products reduces [11].

It is worth remarking that we are mainly interested in the capacity of the FP5 to create and transfer information and nothing can be said about this issue inside each node. Notice that some participants are large institutions or companies with complex organization charts, which may have several projects whose coordination cannot be guaranteed in general. However, our main concern is how to set the means to integrate research, development and innovation efficiently, not to determine if these means are successfully used.

2. Analysis of the data

To characterize the FP5, in this section we compute five important features in any network: degree distribution, shortest path distribution, betweenness centrality, clustering coefficient and the degree–degree correlation. The detailed description of the dataset can be found in the appendix.

2.1. Degree distribution

The probability that a university collaborates with k other universities (i.e. the degree distribution of the universities) decays as a power law, $P(k) \sim k^{-\gamma_U}$ with $\gamma_U = 1.76$. Similarly, companies follow a power law with $\gamma_C = 2.76$. The two distributions can be seen in figure 1, where a log–log scale is used in the plot, providing evidence for the scale-free topology [12] of both networks. The degree distribution of the whole FP5 network is also well approximated by a power law with exponent γ close to 2.1 [13].

Note that the degree distribution of universities is described by a power law with $\gamma_U < 2$, implying that their mean degree grows in time. Indeed the first moment (i.e. mean degree in this

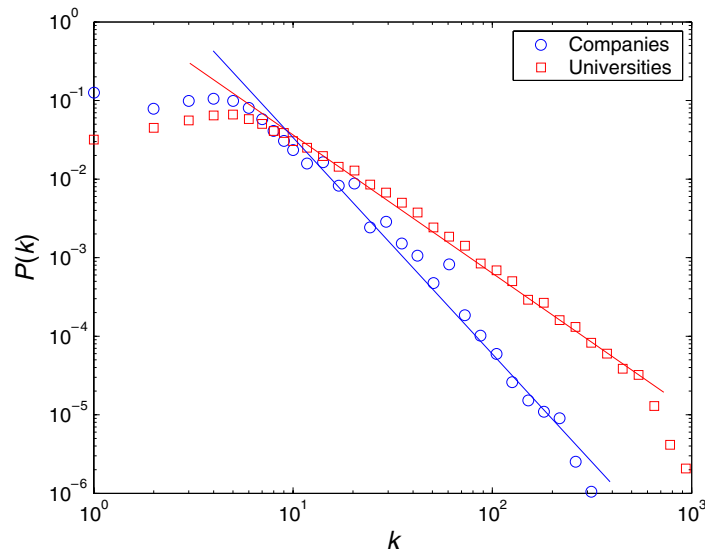


Figure 1. This graph depicts with red squares the probability that a university collaborates with k other universities, that is, its degree distribution. The degree distribution of companies is shown with blue circles. Data were log-binned. We find that both distributions follow a power law tail, $P(k) \sim k^{-\gamma}$, and thus have a scale-free topology, with vertices connecting each other in a heterogeneous manner: most vertices have few connections, but some have a very large degree. The best fit for the straight region of the curves gives $\gamma_U = 1.76 \pm 0.01$ with a correlation coefficient $R = 0.998$ for universities, and $\gamma_C = 2.76 \pm 0.03$ with $R = 0.991$ for companies. However, the fact that universities show $\gamma_U < 2$ whereas companies have $\gamma_C > 2$ implies that the mean degree of universities grows in time but the mean degree of companies does not. This result suggests that some form of synergy encourages the creation of new collaborations mainly between universities, while the network of companies is less dynamic in this respect.

case) of a distribution with a power-law tail diverges when its exponent is less than 2. This result suggests that universities form an accelerated growing network [14], where the total number of edges grows faster than a linear function of the total number of vertices and, consequently, it is verified that $1 < \gamma < 2$.

To elucidate this issue, we computed the average degree $\langle k \rangle$ during several years to check its tendency. Though we only have the data corresponding to 4 years (table 1), they are enough to confirm the existence of an accelerated growth since the average degree is not constant (46% increase for the network of universities in the four year period). But if the collaborations grow faster than proportional to the number of participants, it is because they do not emerge by the mere increase of participants. Not only new participants contribute to increase the number of collaborations, but also the old ones, meaning that some form of synergy exists encouraging the creation of new collaborations between universities.

On the other hand, the average degree of companies also grows (though significantly slower) during the four year span of the dataset (table 1). However, the fact that $\gamma_C > 2$ suggests that this

Table 1. Evolution of universities and companies during the FP5. Here we show the total number of vertices N , the average degree $\langle k \rangle$ and the average clustering coefficient $\langle C \rangle$ during the four years that the FP5 lasted.

Universities–Companies			
Year	N	$\langle k \rangle$	$\langle C \rangle$
1999	3075–4658	17.2–6.2	0.65–0.58
2000	5377–9359	21.9–6.8	0.66–0.53
2001	7355–13905	27.7–7.9	0.67–0.53
2002	8522–16765	31.9–8.2	0.68–0.59

increase should be transient. Therefore, although the creation of collaborations is encouraged (since when the FP5 was finished the mean number of collaborations had risen from 10 to 26 and some participants had surpassed 2500 collaborations) these results reveal that the synergy is more pronounced between universities. In this sense, the FP5 is less effective in improving the network of companies, and universities seem to take more advantage of this opportunity to create new collaborations.

Also noticeable in table 1 is the fact that the number of companies increases faster than the number of universities (72 and 64% increase respectively in the four year period), indicating another difference in the evolution of the two networks.

2.2. Shortest paths

A *path* between two participants is defined as a sequence of edges which links them, the distance between them being the number of edges in the shortest path. Defining the set of participants which can be linked through a path as a connected component, we find that the largest connected component of universities spans 93.7% of the network (7987 vertices) while for companies it is made of 10 801 nodes (64.4%). Hence, while almost all universities are linked in only one component, companies are more fragmented and one third of them fall in other smaller components (actually, the second biggest component contains only 48 participants). This result shows that universities are important to compact the network since the largest connected component of the complete network ($U + C$) comprises 88.7% of the companies and 96.0% of the universities (i.e. 23 055 vertices in total). In addition, the largest distance in the network of universities is 7 and the average distance is $\langle d \rangle = 3.34$ whereas, in the case of companies, the farthest pair is separated by 14 edges and the average distance is $\langle d \rangle = 5.67$.⁴ This can be seen in figure 2 where we plot the distance distribution, $P(d)$ versus d . Hence, also here universities are essential for companies since the largest distance in the entire network is only 8 and the average distance is $\langle d \rangle = 3.14$, which implies that, on average, there are only two intermediaries between two participants.

The average distance is a coarse characteristic though. As a finer measure, it is possible to compute the average distance of a vertex of degree k to all other vertices in the largest component [15]. In figure 3 we plot $\langle d \rangle(k)$ for both networks on a log–linear scale, where the y -axis is $\langle d \rangle(k)$ and the x -axis is $\log k$.

⁴ Both average distances are approximately the value obtained for a random graph [19] with the same number of nodes and average degree. For universities $\langle d \rangle \approx \log N / \log \langle k \rangle = 2.61$ and for companies $\langle d \rangle = 4.62$.

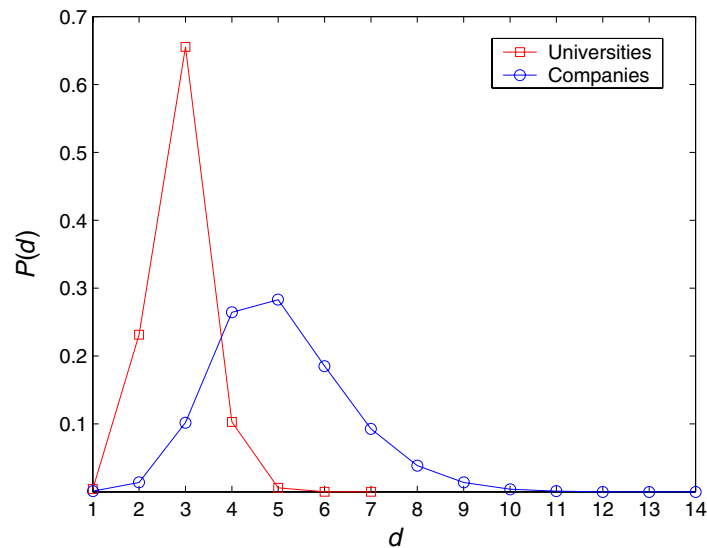


Figure 2. The distribution of shortest paths in the largest connected component of universities (red squares) and companies (blue circles) displays the presence of the small-world effect. The mean value is $\langle d \rangle = 3.34$ for universities and $\langle d \rangle = 5.67$ for companies. Moreover, while the farthest pair of companies has 13 intermediaries, for universities the maximum separation is 7 edges. Therefore, universities are important for companies since, when they cooperate, in the whole FP5 network the largest distance reduces to 8 and the average distance to 3.14.

Therefore, although both networks display the so-called small-world effect [16], there are important differences. The presence of universities eases the flow of information since they are much closer to each other than companies. This could be expected since the main purpose of a company is to satisfy its shareholders, which does not include the spread of information from which competitors can take advantage. But, interestingly, the consequences of this fact go further. When universities are excluded from the projects, companies become isolated despite the fact that universities are only one third of the participants. Companies tend to form clusters, making communication between them difficult (if not impossible) and, consequently, little can be developed or innovated since other results are not available to work with. Thus the natural tendency of companies to protect their findings would end up killing R + D + I. The presence of universities contributes to moderating this.

2.3. Betweenness centrality

To further investigate the interplay between the two kinds of participants, we can also measure the *betweenness centrality* [17] in the FP5. The betweenness σ_m of vertex m measures the extent to which m lies on the paths between other participants. It therefore accounts for the influence of a participant between two other distant participants, relating the local structure and the global topology of the network. It is defined as

$$\sigma_m = \frac{1}{(N-1)(N-2)} \sum_{i,j:i \neq j \neq m} \frac{B(i,m,j)}{B(i,j)},$$

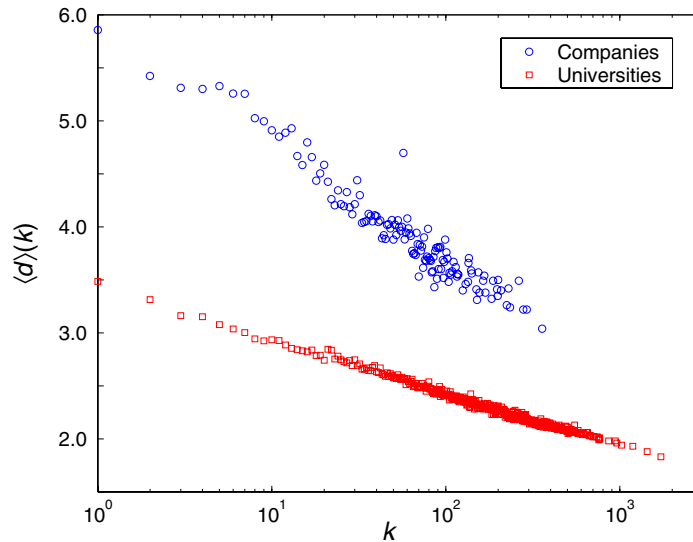


Figure 3. The average distance of a participant with k partners to all other participants in the largest connected component is depicted. Universities are the red squares and companies are the blue circles. The logarithmic dependence can be seen since it is verified that $\langle d \rangle \sim -\beta \log k$ where $\beta_U = 0.503 \pm 0.003$ with $R = 0.994$ for universities and $\beta_C = 1.13 \pm 0.03$ with $R = 0.958$ for companies. The decay is faster (i.e. $\beta_C > \beta_U$) in the network with the larger value of exponent γ (see figure 1), providing empirical evidence for the results of [15]. Note that the lowest degree vertices in the network of universities show a distance to other vertices comparable to the one of the highest degree vertices in the network of companies. Also note that in both networks $\max(k) \approx 2 \min(k)$ as has previously been observed in network models [15].

where $B(i, j)$ is the number of shortest paths between nodes i and j , $B(i, m, j)$ is the number of such shortest paths passing through vertex m , and the sum is taken over all pairs of vertices i and j which do not include m . The pre-factor, where N is the total number of nodes, accounts for normalization, so that $0 \leq \sigma_m \leq 1$.

Since the computation of the betweenness for the whole FP5 is an extremely time-consuming task, we focus our study on one of its subprogrammes: ‘Promotion of innovation and encouragement of small and medium sized enterprises participation’ (SME), which is made up of 195 research institutions and 212 companies (see appendix). Given our ability to split the SME into universities and companies, several different situations are considered. The average betweenness of the SME, taken over all vertices, turns out to be $\langle \sigma \rangle = 5.19 \times 10^{-3}$. Considering only those vertices m which are universities, we find that their average betweenness among all other vertices in the SME is $\langle \sigma_U \rangle = 6.76 \times 10^{-3}$. Likewise, we obtain $\langle \sigma_C \rangle = 3.74 \times 10^{-3}$ for companies.

Now, if we only take into account those shortest paths whose endpoints are companies, the betweenness measures the role universities play in linking companies: $\langle \sigma_{CUC} \rangle = 5.44 \times 10^{-3}$; on the other hand, when the endpoints are universities, the average betweenness of companies is $\langle \sigma_{UCU} \rangle = 2.34 \times 10^{-3}$. Thus, we see that the role universities play between companies is more than twice the one played by companies between universities. Moreover, given that

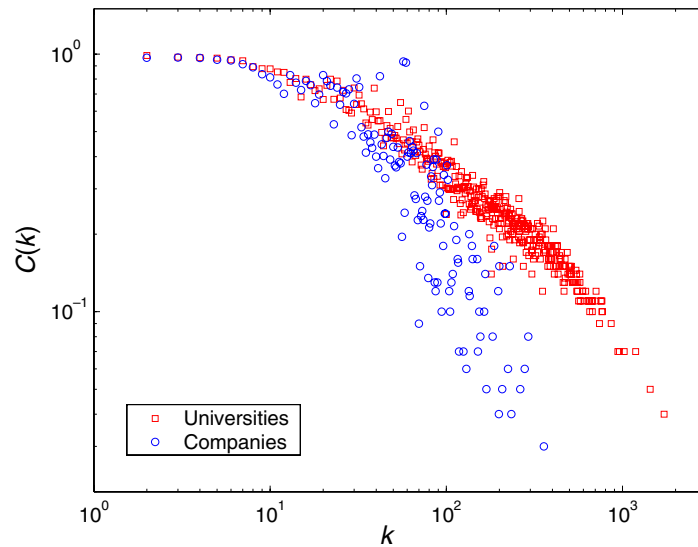


Figure 4. In this graph the clustering coefficient as a function of k is shown. After the initial plateau, where $C(k)$ is approximately constant, it decays as a power law, $C(k) \sim k^{-\alpha}$, where $\alpha_U = 0.54 \pm 0.01$ with $R = 0.97$ for universities (red squares) and $\alpha_C = 1.05 \pm 0.06$ with $R = 0.86$ for companies (blue circles). We therefore conclude that both networks have hierarchical modularity since scale-free and modular networks are degree-independent, whereas hierarchical modularity is characterized by the power-law decay $C(k) \sim k^{-\alpha}$.

$\langle \sigma_U \rangle > \langle \sigma \rangle > \langle \sigma_C \rangle$, we observe again the central function played by research institutions in the FP5 network.

2.4. Clustering coefficient

The *clustering coefficient* of a vertex i is defined as $C_i = 2n_i/[k_i(k_i - 1)]$, where n_i is the number of edges connecting its k_i nearest neighbours. It equals 1 for a participant at the centre of a completely connected cluster, and 0 for a node whose neighbours are not linked at all. Taking the average of the clustering coefficient, we obtain $\langle C \rangle = 0.68$ for universities and $\langle C \rangle = 0.59$ for companies, which are much higher than the average clustering coefficient of a random graph [18] with the same number of nodes and average degree (namely, $\langle C \rangle = \langle k \rangle / N$). Moreover, $\langle C \rangle$ is independent of the number N of participants in both cases (see table 1), in contrast with the prediction of a scale-free model [12] where $\langle C \rangle \sim N^{-0.75}$ [5, 19]. This high and size-independent average clustering coefficient evidences the organization of universities and companies in modules.

However, when we measure the clustering coefficient of a node with k links, $C(k)$, for both networks (figure 4), we find that it decays as a power law for large k . We therefore infer that the two nets have hierarchical modularity, which is characterized by the scaling law $C(k) \sim k^{-\alpha}$, in contrast to some scale-free or modular networks where the clustering coefficient is degree-independent [20].

This result suggests that universities and companies have an inherent self-similar structure [21], being made of many highly connected small modules, which integrate into larger modules,

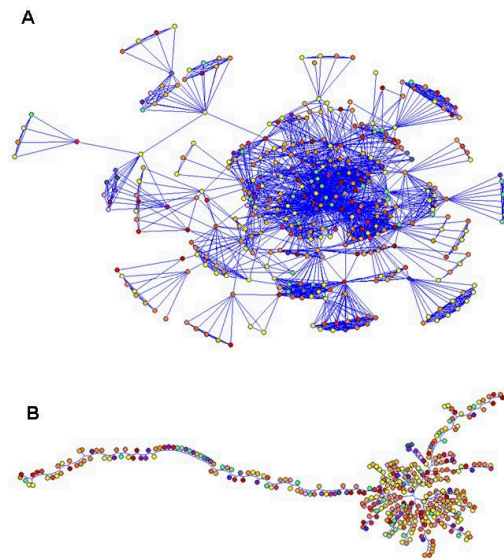


Figure 5. The existence of hierarchical modularity in the networks of universities and companies suggests that they have a self-similar structure. Since projects in the FP are classified into 8 subprogrammes depending on their objectives, we choose, for clarity, to illustrate in (A) this self-similar structure with the smallest one: ‘Promotion of innovation and encouragement of small and medium sized enterprises participation’ (SME). Also, to verify if there is a bias by nationality in the collaborations, we searched for communities reflecting groups of participants collaborating strongly among them. In the networks of universities, companies and both together (even when they are analysed by subprogramme) the result was similar to (B), corresponding to the SME subprogramme. If we colour the nodes according to their nationalities and arrange them in space with a standard algorithm [24], we find that they are all mixed.

which in turn group into even larger modules (figure 5(A)). Actually, we observe that 4333 universities (50.8%) and 10 564 companies (63.5%) have $C_i = 1$, indicating the presence of many totally connected groups. This is due to the fact that most of these entities participate in only one project, having as neighbours other vertices, which in turn are all connected between them by virtue of the participation in the project. Furthermore, given that this result suggests weak geographical constraints [22], we searched for communities in them [23] and found precisely that they were not based on nationality (figure 5(B)), whence, the FP is successfully applying a policy which avoids segregation by nationality.

2.5. Degree–degree correlation

An interesting question is which vertices pair up with which others. It may happen that vertices connect randomly, no matter how different they are. Usually, however, there is a selective linking, i.e. there is some feature which makes the connection more (or less) likely [25]. There is *assortative mixing* when vertices of similar degree tend to be connected, and *disassortative mixing* in the opposite case (i.e. when vertices of high degree tend to connect to vertices of low degree) [26, 27].

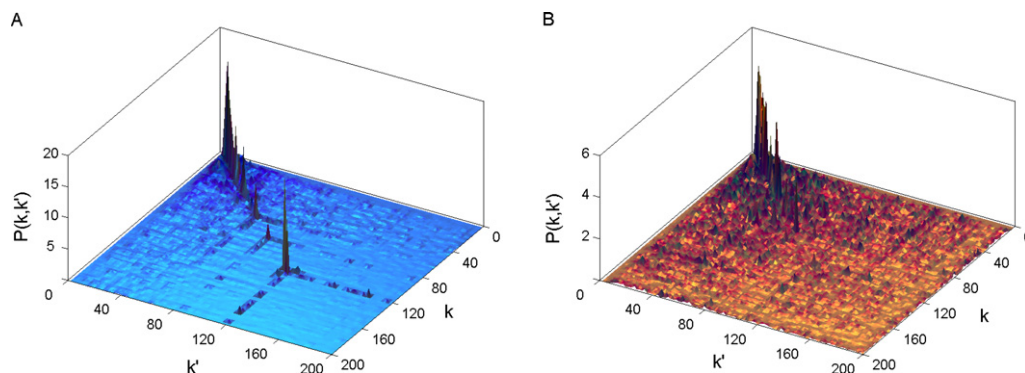


Figure 6. Determination of the mixing through the joint degree–degree distribution. The x - and y -axes represent the degrees k and k' and the z -axis gives the corresponding joint degree–degree probability in per mille. The range is limited from 0 to 200 to illustrate a clearer picture. The joint degree–degree distribution of companies (A) peaks on the line $k = k'$ which implies that the mixing is assortative. Since the number of links held by a participant is related to its size, we infer that companies with similar sizes tend to collaborate more frequently than companies with different sizes. The joint degree–degree distribution of universities (B) is distributed throughout the x – y -plane which suggests that universities do not have assortative mixing and thus choose their collaborators in a less selective manner.

A first approach to elucidate this issue is by means of the *joint degree–degree distribution* $P(k, k')$, which gives us the probability of finding an edge connecting vertices of degree k and k' . We see that for companies the distribution has sharp peaks for $k = k'$ (figure 6(A)). This network thus seems to display assortative mixing, i.e. if one chooses at random a vertex of degree k then, with considerable probability, it will be connected to vertices of degree k . In other words, companies with similar degree tend to collaborate more frequently than companies with different degrees.

Notice that the fact (mentioned in the previous section) that many entities participate in only one project may, by itself, explain these peaks: if the X participants of a certain project have no other projects each of them has degree $X - 1$ and each of their neighbours has degree $X - 1$, giving rise to an assortative trend. On the other hand one can also argue that, when a company has high degree it is due to being involved in many projects. It is then reasonable to assume that nodes with high degree represent large institutions, given that only these can deal with many projects at the same time. That being the case, the observed assortativity means that the spread of information between companies depends on the institution's size. On the contrary, for universities $P(k, k')$ is scattered throughout the plane k – k' (figure 6(B)). While there are still peaks along the line $k = k'$, the presence of many others for $k \neq k'$ is clear, suggesting that universities are less selective in what regards the size of their partners.

It is important to remark, however, that the joint degree–degree distribution requires many observations in order to obtain good statistics. For example, if we focus our analysis in the range $[0, 200]$, we need about 200×200 points, otherwise fluctuations are important and the plot is far from smooth [28]. To avoid this problem, one uses the average degree of the nearest neighbours

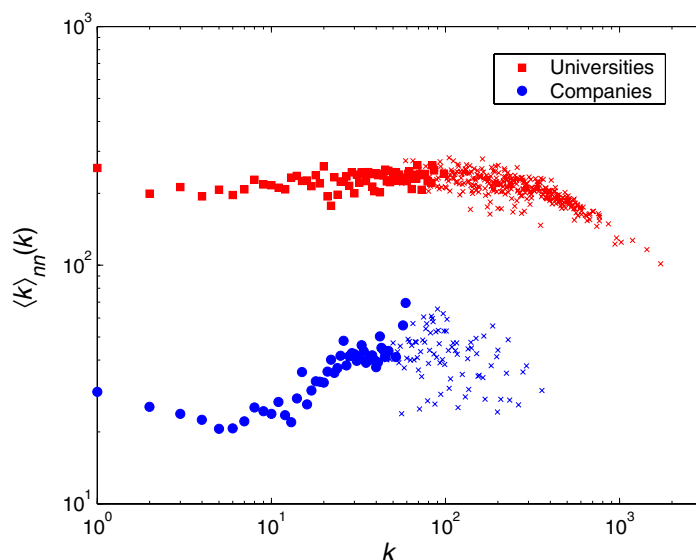


Figure 7. In this plot the average degree of the nearest neighbours of a vertex with k links, $\langle k \rangle_{nn}(k)$, is shown. To mark the proximity to the cutoff, the points obtained from less than 10 observations are plotted as crosses (universities in red and companies in blue) and the remaining points as squares (universities) or circles (companies). In this manner, it can be seen that these points are biased downwards due to the finite size of the network. Then, once focusing our attention on the circles and the squares, we find that companies have assortative mixing, while universities link between them regardless their degrees.

of a vertex of degree k , $\langle k \rangle_{nn}(k)$, which is a coarser but less fluctuating measure. To compute it, we find all participants with k links and take the average degree of all their neighbours. The results are shown in figure 7, and confirm those obtained through the joint degree–degree distributions. To emphasize the presence of the cutoff due to the finite size of the network, the points obtained from less than 10 observations are plotted as crosses (universities in red and companies in blue) and the rest of the points as squares (universities) or circles (companies). Considering only the circles and the squares, we confirm that collaborations between companies are size-dependent (positive slope) whereas those between universities are not (no slope).

It is also interesting to analyse how universities and companies link each other, which can be done as follows. We search for all companies with k links and then compute the average degree of all their neighbouring universities. Note that the former degrees are always calculated in the corresponding network, thus a company with degree k has k neighbour companies, though it may have more links (to universities) in the complete FP5 network. Analogously, we can find all universities with k links to average the degrees of all neighbour companies. The results are depicted in figure 8 where, as before, a log–log scale is used. Again, we plot as squares (universities) or circles (companies) the points obtained from more than 10 observations to identify the region where the tendency is well defined. We find that, while companies link to universities independently of their sizes, universities with high degree tend to collaborate with large companies.

Finally, another way to quantify the mixing in the FP5 is by means of the *assortativity coefficient* [26], which is just the Pearson correlation coefficient of the degrees of connected

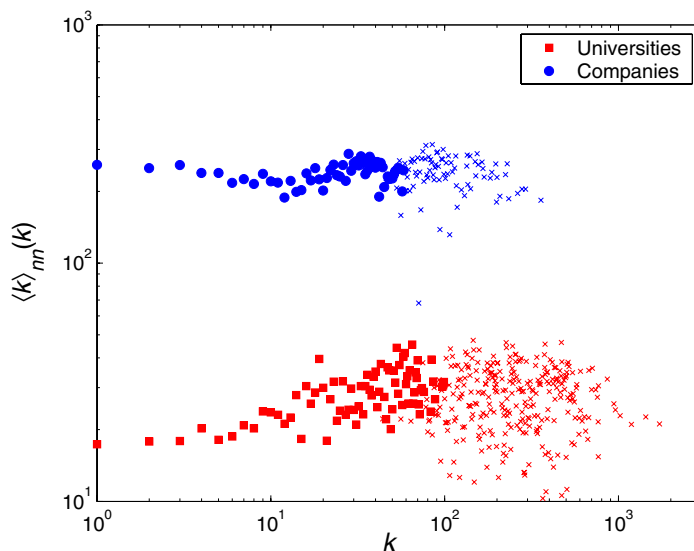


Figure 8. Here we plot the average degree of the nearest companies of a University with k links to other universities (red squares) and the average degree of the nearest universities of a company with k links to other companies (blue circles). As before, if we only consider the circles and the squares, we find that companies link to universities independently of their degrees while universities with high degree collaborate mainly with companies which also have high degree.

vertices. In this case, we obtain what type of mixing takes place in the network by means of a single number instead of a distribution. If e_{jk} is the probability that a randomly chosen edge has vertices with degree j and k at either end, the assortativity coefficient takes the following form:

$$r = \frac{\sum_{jk} jk(e_{jk} - q_j q_k)}{\sum_k k^2 q_k - (\sum_k k q_k)^2},$$

where $q_k = \sum_j e_{jk}$ and $q_j = \sum_k e_{jk}$. This coefficient verifies that $-1 \leq r \leq 1$, being positive when the network is assortative and negative when it is disassortative. We find $r_C = 0.13$ for the network of companies and $r_U = 0.06$ for universities, corroborating an assortative trend usual in social networks [27].

Therefore, companies and universities differ in the way they establish collaborations. Companies are organized hierarchically, where positions in that hierarchy are related to the size: the assortative trend in the network of companies suggests that large corporations are reluctant to choose small companies as partners. Between universities, however, size is not important and it is common to find a large institution collaborating with a small one. But if we analyse which partners universities choose among companies, we find that large institutions in universities prefer working with large companies. On the contrary, companies select their collaborating universities regardless of their sizes. We can then conclude that large companies indeed play a leading role in the FP5 while universities play the role of bridges between participants which are separated in the hierarchical structure of companies.

3. Conclusion

We have presented here a study of the interplay between research and industry in the scope of the FP5. Using network theory methods, we perform several measures that allow us to quantify the features of this relationship and assess their potential improvements. Naturally, the FP5 network does not include all interactions between universities and industry (such as the recruitment of graduates by companies, the transfer of knowledge through scientific and technical literature or industry conferences). Furthermore, as already mentioned in the introduction, it also neglects the fact that internal connections in an institution (e.g. between different departments) may be absent, which would mean that a node in the studied network would split into disconnected nodes. While these issues may significantly influence the flow of information in the network, addressing all of them requires information that is beyond the reach of most researchers at this point. The presented analysis thus represents a starting point for a quantitative understanding of the university–industry interplay network. It is possible, however, to foresee advances in these directions, given the increasing availability of information on how institutions self-organize.

The results point to the central function played by universities in the FP5 network in reducing the distance between research and applications. Indeed, we show that universities play a crucial role in connecting the network of companies, which would otherwise be separated in many small clusters. While the network of universities is well integrated and established in accordance to what is observed for other social networks, the same does not seem to apply for the companies network, mainly due to its relatively small largest connected component. Competition is probably the origin of this effect, which is moderated by the presence of universities. It seems reasonable, then, to conclude that special attention should be devoted to company–company collaborations. Supporting this, there is also the fact that new collaborations arise at a higher rate between universities.

Our observations suggest in addition that companies and universities establish collaborations differently: while companies seem to exhibit a hierarchical structure in terms of their size, universities are less selective in their collaborations. We also observed that both networks display hierarchical modularity and that communities in the FP5 network are not nation-based. The FP appears then to mix all nationalities of the EU, thus reaching one of its main goals: to promote the transfer of knowledge throughout Europe.

Acknowledgments

JAA, LL, and MAFS acknowledge financial support from the Spanish Ministry of Science and Technology under project number BFM2003-03081 and project FIS2006-08525. JGO acknowledges financial support from FCT Portugal (SFRH/BD/14168/2003). JFFM was partially supported by projects POCTI (FAT/46241/2002 and MAT/46176/2002) and project DYSONET.

Appendix. Classification of participants into companies and universities

The FP sets out the priorities for the EU's research and technological development. These priorities are defined following a set of criteria which pursue an increase of the industrial competitiveness and the quality of life for European citizens. A fact which shows the effort

made by the EU to promote this global policy for knowledge is the budget devoted to these programmes. For example, the FP5 (1998–2002) was implemented by means of 13 700 million euros and the FP6 (2002–2006) has assigned a budget of 17 883 million euros.

All projects in the FP5 are organized in eight specific programmes which can be classified as follows. There are five focused Thematic Programmes implementing research, technological development and demonstration activities.

1. QOL: quality of life and management of living resources (2524 projects).
2. IST: user-friendly information society (2382 projects).
3. GROWTH: competitive and sustainable growth (2014 projects).
4. EESD: energy, environment and sustainable development (1772 projects).
5. NUKE: research and training in the field of nuclear energy (1032 projects).

And there are three horizontal programmes to cover the common needs across all research areas.

1. INCO: confirming the international role of community research (1034 projects).
2. SME: promotion of innovation and encouragement of small and medium enterprises participation (142 projects).
3. HPOT: improving human research potential and the socio-economic knowledge base (4876 projects).

The data to analyse the FP5 as a complex network were obtained from the web pages of CORDIS [10] with a robot implemented in Perl. The result was a database with 15 776 records as follows:

Programme | Year | Participant1 - Nation - Dedication | Participant2 - Nation - Dedication | . . .

The first field refers to the specific programme to which the project belongs and the second field informs us about the year in which it started. The following fields are the participants in the project with their corresponding nationality and dedication ('research', 'education', 'industry' . . .). We then have a bipartite graph [5, 14] since there are two kinds of vertices (participants and projects) and each edge links a participant with a project. To obtain the graph with 25 287 participants (nodes) and 329 636 collaborations (edges) used throughout the text, we have only to project it onto the participants.

The names of the participants were not free of typos since we collected them as they were in the web. The consequence of this fact was that sometimes the same participant appeared in two projects with different names and, consequently, it was recorded twice in the data. For instance, 'François Company of Something, Ltd' and 'Francois Company of SOMETHING LTD' would be recorded as different. To avoid these duplications, we used a parser covering many possibilities which could lead to false entries. Nevertheless, despite our efforts, not all duplications have been eliminated. However, after a visual inspection of the data, we estimate that the error is below 10%.

To split the participants into universities and companies, we considered the organization type reported in the project. This information is encoded in the field 'dedication', where we found

11 levels: ‘commission external service’, ‘commission service’, ‘consultancy’, ‘education’, ‘industry’, ‘non commercial’, ‘not available’, ‘other’, ‘research’, ‘technology transfer’ and ⟨void⟩.

The level ‘not available’ means that the FP itself was not able to obtain the information and this absence is shown in this manner. In addition, the level ⟨void⟩ means that no information at all is given, i.e. our robot found nothing (not even ‘not available’).

The first step to define only two groups was to reduce the number of levels in ‘dedication’. We found that eight levels could be merged to define a new one, called ‘non companies’. It was not homogeneous since we found consultancies, universities, hospitals, institutes, laboratories, observatories, museums, technological parks, even cities. However, they all were participants involved in some type of research for whom results do not necessarily return income. This new level was, basically, the union of ‘research’ and ‘education’ since the other six levels appeared few times in the data: ‘commission external service’ (4 records), ‘commission service’ (8 records), ‘consultancy’ (49 records), ‘non commercial’ (389 records), ‘technology transfer’ (1 record) and ⟨void⟩ (1 record). The record with ⟨void⟩ was identified as ‘non company’ by direct inspection.

Therefore, all records could be classified in one of the following levels: ‘non companies’ (41 317), ‘industry’ (6447), ‘other’ (17 588) and ‘not available’ (12 346). The total number of records (77 698) is larger than the number of participants (25 287) since many of them collaborate in several projects. Then, it was necessary to verify if repeated records were always classified in the same level of ‘dedication’.

We found that many participants were classified in different levels, thus we had to define a set of rules which eliminated this ambiguity. Hence, the following step was to study each level to understand their composition. For every level, we chose 100 records randomly to check by direct inspection their dedication. The result was that all selected records in ‘industry’ were companies, none in ‘non companies’, 95 in ‘other’ and 55 in ‘not available’.

With the former information, we proceeded as follows. We first defined for each participant a vector $D = \{\text{‘non companies’}, \text{‘industry’}, \text{‘other’}, \text{‘not available’}\}$, where the components are the number of times that it is classified in that level. For instance, $D = \{17, 0, 8, 4\}$ means that the participant appears 17 times as ‘non company’, 8 as ‘other’ and 4 as ‘not available’. Then, we decided that vectors in the form $\{a, 0, 0, 0\}$ or $\{a, 0, 0, d\}$ were universities and vectors in the form $\{0, b, c, d\}$, $\{0, b, c, 0\}$, $\{0, b, 0, d\}$ and $\{0, b, 0, 0\}$ were companies. With only these sensible rules, we managed to classify 22 001 participants (87%).

In order to confirm this result and to classify the remaining 3286 entities, we defined a filter based in keywords relative to the universities group, such as ‘univer’, ‘schule’, ‘laborato’ When we focused our attention in the group of 22 001 participants classified using ‘dedication’, we found that those classified as universities according to the filter were also universities according to ‘dedication’. Since the filter was a completely different manner of splitting the dataset, we could use it for the rest of the entries. Note that we only believed the result of the filter if it was university, not if the result was company. This is reasonable since the filter was designed to identify terms related to universities, not to companies.

By means of the filter we classified all participants but 309. To place these entities, we paid attention to which value was higher: ‘non companies’ or ‘industry’, independently of the other two values. If the value ‘non companies’ was higher, it was a university, otherwise it was a company.

References

- [1] Amidon D 2004 *Baltic Dynamics Conf.* Riga
- [2] Branscomb L M, Kodama F and Florida R L 1999 *Industrializing Knowledge: University-Industry Linkages in Japan and the United States* (Cambridge: MIT Press)
- [3] Caloghirou Y, Tsakanikas A and Vonortas N S 2001 University-industry cooperation in the context of the European Framework Programmes *J. Technol. Transfer* **26** 153
- [4] Meyer-Krahmer F and Schmoch U 1998 Science-based technologies: university-industry interactions in four fields *Res. Policy* **27** 835–51
- [5] Albert R and Barabási A-L 2002 Statistical mechanics of complex networks *Rev. Mod. Phys.* **74** 47–97
- [6] Guimerà R and Amaral L A N 2005 Functional cartography of complex metabolic networks *Nature* **433** 895–900
- [7] Jeong H, Tombor B, Albert R, Oltvai Z and Barabási A-L 2000 The large-scale organization of metabolic networks *Nature* **407** 651–4
- [8] Dorogovtsev S N and Mendes J F F 2003 *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford: Oxford University Press)
- [9] Balthrop J, Forrest S, Newman M E J and Williamson M M 2004 Technological networks and the spread of computer viruses *Science* **304** 527–9
- [10] Community research and development information service online at <http://cordis.europa.eu>
- [11] Wigzell H 2002 Framework programmes evolve *Science* **295** 443–5
- [12] Barabási A-L and Albert R 1999 Emergence of scaling in random networks *Science* **286** 509–11
- [13] Almendral J A, Oliveira J G, López L, Mendes J F F and Sanjuán M A F to be published
- [14] Dorogovtsev S N and Mendes J F F 2002 Evolution of networks *Adv. Phys.* **51** 1079–145
- [15] Dorogovtsev S N, Mendes J F F and Oliveira J G 2006 Degree-dependent intervertex separation in complex networks *Phys. Rev. E* **73** 056122
- [16] Strogatz S H 2001 Exploring complex networks *Nature* **410** 268–76
- [17] Freeman L C 1977 A set of measures of centrality based on betweenness *Sociometry* **40** 35–41
- [18] Bollobas B 1985 *Random graphs* (London: Academic)
- [19] Bollobás B and Riordan O M 2003 Mathematical results on scale-free random graphs *Handbook of Graphs and Networks: From the Genome to the Internet* ed S Bornholdt and H G Schuster (Berlin: Wiley) pp 1–34
- [20] Ravasz E, Somera A L, Mongru D A, Oltvai Z N and Barabási A-L 2002 Hierarchical organization of modularity in metabolic networks *Science* **297** 1551–5
- [21] Chaoming S, Havlin S and Makse H E 2005 Self-similarity of complex networks *Nature* **433** 392–5
- [22] Ravasz E and Barabási A-L 2003 Hierarchical organization in complex networks *Phys. Rev. E* **67** 026112
- [23] Newman M E J 2004 Fast algorithm for detecting community structure in networks *Phys. Rev. E* **69** 066133
- [24] The algorithm package PAJEK can be found online at <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>
- [25] Newman M E J 2003 The structure and function of complex networks *SIAM Rev.* **45** 167–256
- [26] Newman M E J 2002 Assortative mixing in networks *Phys. Rev. Lett.* **89** 208701
- [27] Newman M E J 2003 Mixing patterns in networks *Phys. Rev. E* **67** 026126
- [28] Boguñá M, Pastor-Satorras R and Vespignani A 2004 Cut-offs and finite size effects in scale-free networks *Eur. Phys. J. B* **38** 205–10