EPJ B
Condensed Matter and Complex Systems
EPJ.org
your physics journal

Eur. Phys. J. B (2014) 87: 292 DOI: 10.1140/epjb/e2014-50450-4

Network of participants in European research: accepted versus rejected proposals

Maria Tsouchnika and Panos Argyrakis

edp sciences
Società Italiana di Fisica
Springer

Eur. Phys. J. B (2014) 87: 292
DOI: 10.1140/epjb/e2014-50450-4

THE EUROPEAN PHYSICAL JOURNAL B

Regular Article

Network of participants in European research: accepted versus rejected proposals

Maria Tsouchnika and Panos Argyrakis[a]

Department of Physics, University of Thessaloniki, 54124 Thessaloniki, Greece

Received 2 July 2014 / Received in final form 10 October 2014
Published online 8 December 2014 –

Abstract. We investigate the network formed by the collaboration of researchers seeking funding by the European Commission by submitting research proposals. Institutions are network nodes and collaborations are links between the nodes. We constructed one network for the accepted proposals and one for the rejected ones, in order to look for any structural differences between them. To this end, first, we compare the size of the largest connected components and the resulting degree distributions. The latter show notable difference only in the region of relatively small degrees. We calculate the assortative mixing by participant type, i.e. a property which indicates whether the participant is a university/research institute, a company (non-profit included), or undefined. By aggregating the data of both networks into three geographical scales (city, region, country), we compare the degree assortativity and average node weight, in all scales. With respect to these two features the networks display similar behaviour. Finally, we compare a series of centrality measures and the Minimum Spanning Trees, at the country scale, to assess the relative performance of the countries. We find that five countries, France, Germany, the United Kingdom, Spain and Italy, play a central role in both networks, however, their relative significance is not the same.

1 Introduction

Amongst the goals of the European Commission (EC) is to enhance research and technological advances in its member countries, by strongly encouraging collaboration between them. The Framework Programmes (FP), which promote and fund collaboration projects in which European research institutions and companies participate, is one way of achieving this. There has been a series of such FPs over the past three decades, each program lasting for several years. The recent ones include FP5 (1999–2003), FP6 (2003–2007) and FP7 (2007–2013). While the bulk of support and number of partners come from the EC 28 member countries, practically every country in the world may participate, as it is shown by the total of 169 countries involved in recent projects.

As previously mentioned, one of the main goals of the Framework Programmes is to boost research and technological achievements [1–3]. Several publications proceeded to assess the FP effectiveness, in terms of different, yet, closely related objectives, which contribute to this main goal. The two principal points that were examined were (a) whether there is sufficient collaboration between academia and industry to foster linking research and innovation [1,3–6]; and (b) whether the participation of member countries is balanced, in other words whether there is a fair involvement of both the advanced countries/regions, as well as those that fall behind, ultimately aiming for social, economic and technological convergence [4–7].

[a] e-mail: panos@auth.gr

Several references used statistical methods [1,4,7], as well as network analysis [5,6], to address these queries. They found that the FPs are successful in some aspects, while more effort is needed in others. By comparing the network of companies participating in FPs to that of the universities, they examined the extent to which several variables such as, geographical distance, technological proximity, language barriers, etc., enhance or inhibit the collaboration between countries/regions [4–6]. The hindering effect of the technological distance is the most prominent, for both cases. All these factors appear to have a greater impact on the structure of the industrial network rather than that of academic research. In fact, the research network appears to be serving as a buffer between the companies and reduces the distance between academic research and industry by being the backbone of the network [5]. All evidence suggests that the room for improving is greater in the industrial sector than that of research [5,6]. They also found that during the period 1984–1991 four clusters of neighbouring countries/regions were formed, because of the fact that ties between them are favoured. These four clusters were reduced to three during 1992–1998 [4]. In a later publication, however, that surveyed FP5 (1999–2003), it was found that the observed FP network communities reveal an adequate mixing of nationalities, suggesting that country border effects are not as significant as they had been in the past [5].

In the present work we follow the network approach as in references [5,6]. Within this frame of reference, one constructs the network of all partners, its nodes denoting the partner (research institution, university, company, etc.), and the links between nodes denoting the existence of a proposal/collaborative project between them. Information of interest then includes the size of projects (meaning the number of participating nodes), the type of consortium formed (small focused projects between 2 or 3 partners, versus large networks of dozens of partners on a general theme) the cities, regions, countries and continents participating in the projects, identifying important nodes in the network and other pertinent information. We analyse data of the seventh framework programme (FP7) using network analysis, this time including all submitted proposals in the Call of Interest. Previous work focused only on the proposals that were accepted and subsequently implemented as projects whereas we also include those that were rejected and never carried out. We construct two networks of participants, one for those involved in the accepted proposals and one for those in the rejected ones, and then aggregate the data into three geographical scales (city, region, country) thus constructing six new networks, three for each case. The motive behind this idea is to see whether there are any differences between the two networks formed from these two different sets. Specifically, we compare the two networks in terms of (a) the degree, i.e. the number of collaborators, of the participants, (b) the extent of collaboration between academic and industrial participants, (c) the amount of collaborations between large-scale and small scale participants, regardless of type (academic or industrial) and (d) the most significant/influential countries and their relative importance. For this purpose, we compare the structure of the two networks in all scales by examining several structural features. For the country scale in particular, the comparison is made by assessing five network properties (a series of centrality indices and MST) for both cases.

2 Data description

The data we use were provided by the European Commission, Office of Statistics, in anonymised form. Two databases were provided, in Access format, one for all submitted proposals and one for the accepted proposals, which eventually led to signed contracts and were carried out as projects. Each database contains detailed information about the projects/proposals and their corresponding participants. Of all participant characteristics provided, the only piece of information that can distinguish one from another is its participant identification code (PIC), which is a nine-digit number, rather than the name of the Institute, which might be different in several occasions, due to different language used, epithet of the Institute, etc. About 17% of the proposals database entries had no PIC number and were not included in the subsequent analysis of the participants network, whereas, all – or nearly all – records included the information needed for the networks of the other three geographical scales (cities, regions, countries). In order to construct the data set of rejected proposals, which was not given, we simply removed all the records of the accepted proposals from the data set of all proposals. The remaining proposals are the ones that were rejected. Just as previously, about 18% of records of this new data set is lacking a PIC. Table 1 shows the number of participants, cities, regions and countries involved in FP7. It shows the entries that we subsequently use, in all scales, after removing the ones with missing records. Obviously, since a particular participant may be involved in more than one proposal, the sum of the number of participants in accepted and rejected proposals is not equal to the total number of participants in all scales.

Table 1. Number of participants, cities, regions and countries involved in FP7.

	Participants	Cities	Regions	Countries
Proposals	52823	20200	1771	209
Accepted	24396	7479	1441	169
Rejected	47655	18761	1757	207

3 Results

3.1 Size of largest connected component and degree probability function

Our goal is to compare the networks of collaborations formed by the accepted proposals to those formed by the rejected ones. To this end, we first construct the two corresponding networks of participants (universities, research institutions, companies, etc.) involved in the FP7 accepted and rejected proposals, following this procedure: if two given participants are found to have joined a proposal together, we draw a link between them. A weight of one is assigned by default to this link. Each time that a given pair of participants is found to be involved in another proposal, the weight of their link is increased by one. This results in a weighted network, with its nodes representing the project participants, and its links the existence of at least one proposal/collaboration between them. The weight of its links reflects the strength of this collaboration. Self-loops are discarded during this constructing procedure. We find that, by and large, the largest connected component in both cases comprises the majority of the nodes, specifically 99.1% for the accepted case and 98% for the rejected one.

The maximum and average node degree for the largest connected component of the two networks of participants are shown in Table 2. The normalized degree probability distributions of the two networks are shown in Figure 1a. The distributions are qualitatively similar to each other and also similar to the corresponding distributions of the FP5 and FP6 accepted proposals, found in previous studies [5,6]. Specifically, they are highly right-skewed, heavy-tailed distributions, the shape of which, is indicative of a power law behaviour. At large k values we observe a considerable increase at the noise level, as expected. The only observable difference between the distributions in Figure 1a is that, for degrees ranging from $k = 1$ up to about $k = 10$, the normalized number of participants in rejected proposals is larger than that in the accepted

Table 2. Basic properties of the largest connected component of the networks of collaborations in FP7.

	Participants	Cities	Regions	Countries
Number of nodes (acc.)	24 181	7459	1440	169
Number of nodes (rej.)	46 567	18 451	1753	206
Number of links (acc.)	478 274	214 755	105 050	4538
Number of links (rej.)	1 018 721	608 973	236 061	7325
$\langle k \rangle$ (accepted)	39.2	57.4	145.8	53.7
$\langle k \rangle$ (rejected)	42.8	64.9	268.7	70.8
k_{max} (accepted)	5292	3354	1075	155
k_{max} (rejected)	10 769	6840	1450	192

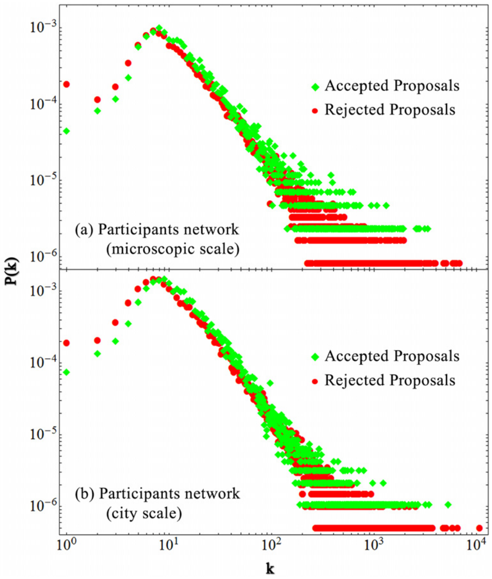

Fig. 1. Degree probability distribution function of (a) the network of participants of the FP7 accepted (rhombi) and rejected (circles) proposals and (b) the same two networks, with the data aggregated in the city geographical scale.

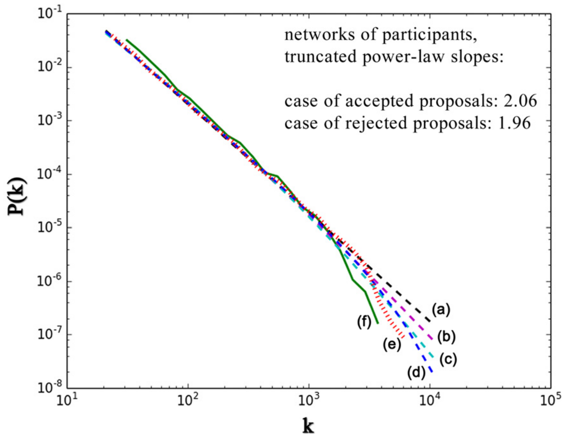

Fig. 2. Degree probability distribution function of the network of participants of the FP7 accepted (solid line (f)) and rejected (dotted line (e)) proposals, using logarithmic binning. Probable fits for the rejected case, in dashed lines: (a) power law, (b) lognormal, (c) stretched exponential, and (d) truncated power law.

ones. This implies that a proposal that consists of participants with relatively small degree is more likely to be rejected. Assuming that low-degree participants mostly represent small-scale institutes in the two networks, a plausible explanation for this finding could be that a group of small-scale institutes alone would strain to meet the criteria of a successful proposal. Therefore, such a consortium could presumably benefit from the addition of one or more larger-scale institutes, which are likely to be the high-degree nodes in the networks.

For values of $k > 10$, the shape of the curves of Figure 1a is indicative of a power law behaviour. Therefore, it seems reasonable to assess how close they are to a probability density function (PDF) of a power-law (scale-free) network. We use the method of maximum likelihood, proposed in reference [8] and the Python powerlaw package presented in reference [9], to test for five probable fits (power-law, truncated power-law, exponential, stretched exponential and lognormal). Using log likelihood ratio tests, we find that of all five the truncated power-law distribution is the most likely fit, in both the accepted and rejected case. The slope of the truncated power-law fit is 2.06 and 1.96, respectively. In Figure 2 we plot the two distributions of Figure 1a, using logarithmic binning for visual clarity, along with four of the five candidate fits for the rejected case, excluding the exponential one, which was rejected as a clearly poor fit. For the range $k = 10$ to $k = 1000$ there is no notable difference in Figure 2 between the distributions of the two networks. All five trial fits fail to capture the behaviour of the tail, i.e. for degrees larger than $k = 1000$, of the data derived distribution. Notice, however, that there are very few nodes with such high degree, implying a large noise level and thus rendering the fitting problem intractable.

3.2 Assortative mixing by participant type and by degree and average node weight

In our databases, each node in the networks of participants can be assigned to three categories according to whether it is (a) an institution affiliated with education/research (university, research facility, etc.), (b) a company (nonprofit included) and (c) undefined. This property is called the participant type. To examine the assortative mixing in the two networks by the participant type we evaluate the corresponding assortativity

Table 3. Degree assortativity coefficient for the network of participants, in all four scales (participants, city, region, country), in both the accepted and rejected case.

Scale	Accepted degree assortativity r	error σ_r	Rejected degree assortativity r	error σ_r
Participant	−0.1018	0.0008	−0.1145	0.0004
City	−0.2608	0.0011	−0.2955	0.0008
Region	−0.3177	0.0027	−0.3475	0.0021
Country	−0.2914	0.0148	−0.3543	0.0126

coefficient r [10]. In an undirected network, r is defined as:

$$r = \frac{\sum_i \left(e_{ii} - \sum_i a_i^2\right)}{1 - \sum_i a_i^2} \tag{1}$$

where e_{ii} is the fraction of edges that run between nodes of the same category and a_i is the fraction of ends of edges attached to nodes of category i. r is positive when there is a tendency of nodes of the same category to be connected to each other (assortative mixing) and negative when the opposite is true (disassortative mixing). The value of r lies in the range $-1 \leq r \leq 1$. $r = 1$ indicates perfect assortativity, $r = -1$ indicates the opposite and $r = 0$ indicates no assortativity. By using equation (1) assortativity coefficient for the participant type for the network of participants involved in the accepted and rejected proposals is $r_{type} = 0.1321$ and $r_{type} = 0.1625$, respectively. Their corresponding expected statistical errors are $\sigma_{r_{type}} = 0.0008$ and $\sigma_{r_{type}} = 0.0005$. We use the jacknife method [11] to calculate the errors, as suggested in reference [10]. We conclude that both networks are assortative, indicating that there is a non-trivial tendency of companies and educational/research facilities to collaborate with participants of the same type. Of the two networks, the one formed by the rejected proposals is more assortative by participant type. A plausible explanation for this fact is that proposals that feature collaborations between academic and industrial participants are favoured.

To attain a more comprehensive picture of the networks, we aggregate the data into the three geographical scales available in the FP7 proposals databases, namely cities, regions and countries and repeated the above procedure, for both the FP7 accepted and rejected proposals, effectively constructing six new networks, two for each scale. The maximum and average node degree for the total of the eight networks in the four scales (participants, cities, regions, countries), are shown in Table 2. The degree probability distributions, in the city scale, are shown in Figure 1b. Both distributions are again right-skewed and heavy tailed, exhibiting a power-law-like behaviour, similar to Figure 1a. In the region and country scale, the corresponding degree probability distributions (image not shown) exhibit a large amount of noise and any possible power-law-like behaviour is blurred.

Using the notion of assortative mixing by degree we measure the probability that two nodes of similar degrees will be linked, in all four scales of the network. This is quantified by the assortativity coefficient, r, which is an example of a Pearson correlation coefficient, similarly to the assortativity coefficient by an enumerative node property [10,12]. In an undirected network, it is defined as:

$$r = \frac{\sum_{jk} jk(e_{jk} - q_k^2)}{\sigma_q^2} \tag{2}$$

where e_{jk} is the fraction of edges that connect a node of degree j to one of degree k, $q_k(=q_j)$ is the fraction of ends of edges attached to nodes of degree k and σ_q is the standard deviation of the distribution q_k. r is positive when there is a tendency of nodes with similar degree to be connected to each other (assortative mixing by degree) and negative when high-degree nodes preferably connect to low-degree ones (disassortative mixing by degree). The value of r lies in the range $-1 \leq r \leq 1$. $r = 1$ indicates perfect assortativity, $r = -1$ indicates perfect disassortativity and $r = 0$ indicates no assortativity (random mixing). Table 3 shows the assortativity coefficient of the two networks of participants involved in the accepted and rejected proposals, respectively, in all four scales, along with the expected statistical errors of these figures. Again, we use the jacknife method [11] to calculate the errors, as suggested in reference [10]. We find that both networks are disassortative by degree in all four scales. This means that participants with small degree mostly end up having high-degree neighbours, in both networks. The network formed by the rejected proposals is more disassortative than that of the accepted, in all four scales. The highest difference is observed in the country scale. The more macroscopic the scale is, the more disassortative both networks become, with the exception of the country scale for the accepted proposals network. The fact that both networks are disassortative is an interesting result on its own, as social networks are known to exhibit assortative mixing by node degree [10]. This finding implies that both networks of the collaborations formed by the FP7 accepted/rejected proposals are structurally different than other social networks. The assortative networks display a dense core consisting of high-degree nodes, with the lower-degree nodes surrounding it, while the disassortative networks, like the two FP7 networks, exhibit star-like features formed because of the tendency of high-degree nodes to connect with lower-degree ones [13]. The reason behind this structural discord between the FP7 networks and other social networks may lay in the circumstances governing the formation of the various collaborating groups intending to carry out a specific project. First of all, it is only reasonable to assume that the small-degree participants mostly represent the relatively small institutes (universities, companies, etc.), while the high-degree represent the major ones, since the latter are in a position to undertake more projects and form more collaborations. The smaller institutes (universities, companies, etc.) do not always have the necessary facilities, equipment, technological knowledge, etc. or their human resources is simply not sufficient for a certain project, while the larger-scale institutes are more likely to be able to provide for all these requirements. Therefore, many small-scale institutes turn to the large-scale institutes to benefit from sharing and collaborating. Moreover, in some cases, the Call for Proposal imposes a collaboration between two specific countries of which one may be a cohesion country which is more likely to have small-scale institutes and the other a more advanced one, having many larger-scale institutes. The different structural features that the disassortative FP7 networks exhibit, in contrast with other social networks that are assortative, harbour two additional implications [10]. The first is that, as a consequence, they are both more susceptible to the removal of high-degree nodes. This is because, in the case of degree disassortativity, the high-degree nodes are spread all over the network, therefore by attacking them one effectively attacks every part of the network. In other words, if for any reason, many of the large-scale institutes left the network simultaneously, it would break down to isolated components and collapse. This means that the large-scale participants are essential to the structure and function of the FP7 network. The second implication is that epidemics in networks with that kind of topology would span to a larger portion of the population than in a similar assortative networks. This could mean that the structure of the two networks is preferable for the spreading of ideas, trends, knowledge, technological advances, etc., which, in this case, is a rather desirable property. Therefore, both networks are more efficient in that way than an assortative network would be, although the network of the rejected proposals case would be more efficient than that of the accepted one.

Next, we calculate the average node weight, $\langle w \rangle$, of every node, i.e. the sum of the weights of all its connections, divided by its degree, for all four network scales, for both the accepted and rejected proposals case. The outcome of this procedure, plotted against the degree of each node, is shown in Figure 3. It is evident that, the behaviour of the two networks is qualitatively comparable, in all respective four scales. Specifically, while in both the accepted and rejected case, in the participant scale, there is no remarkable variation of the average node weight with respect to the node degree, when we aggregate the data into the three geographic scales, there emerges a propensity for the average weight to increase with the degree. The more macroscopic the scale into which we aggregate is, the more intense this phenomenon becomes. These results are in agreement to those of previous studies [6].

3.3 Centrality indices and MST

Directing our attention to the country scale of the two different cases, we set out to compare the two networks in a more detailed way, by using a variety of centrality indices and Minimum Spanning Trees (MST). Each of these

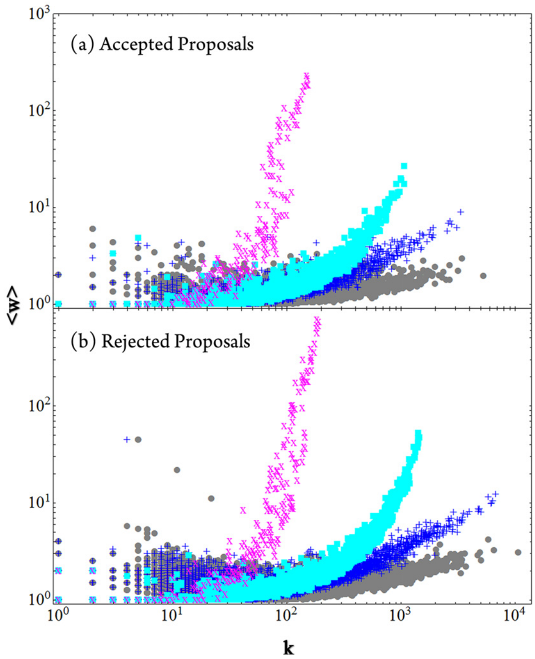

Fig. 3. Average node weight vs. node degree k for the network of participants in microscopic scale (circles), city scale (plus signs), region scale (squares) and country scale (crosses), for (a) the accepted proposals and (b) the rejected proposals.

network properties alone cannot indicate the importance of a certain node, but when considered in combination they help reveal a clearer picture. We start with the simplest of the centrality indices, namely the degree centrality [14] which is simply the degree of a node. The eigenvector centrality [14–16], which we subsequently consider, is very similar to the degree centrality but takes into account that the neighbours of a node are not all equally important. A node scores higher in eigenvector centrality as the relative importance of its neighbours increases. Next, we calculate two indices based on the notion of a shortest path, the closeness [17] and betweenness centrality [18,19]. Closeness centrality is a measure of the distance between two nodes of the network. Betweenness centrality is a measure of how significant is the presence of a certain node in the network in maintaining the connection between two other distant nodes. For the calculation of betweenness centrality, we use the algorithm presented in reference [20].

The four indices mentioned so far do not take into consideration the weights of the links between the nodes of the network. As mentioned above, at the country scale, the average weight has a tendency to increase with the node degree. Nonetheless, there are cases, in which, between two nodes with unequal degrees, the one with the lowest degree has the highest average weight of the two. The presence of link weights perplexes the structure of the network and potentially alters the sense in which a node is important or influences others. Therefore, it is useful to assess the importance of a node with a tool that also takes into account the link weights. For that purpose we employed two more network analysis tools, a variation

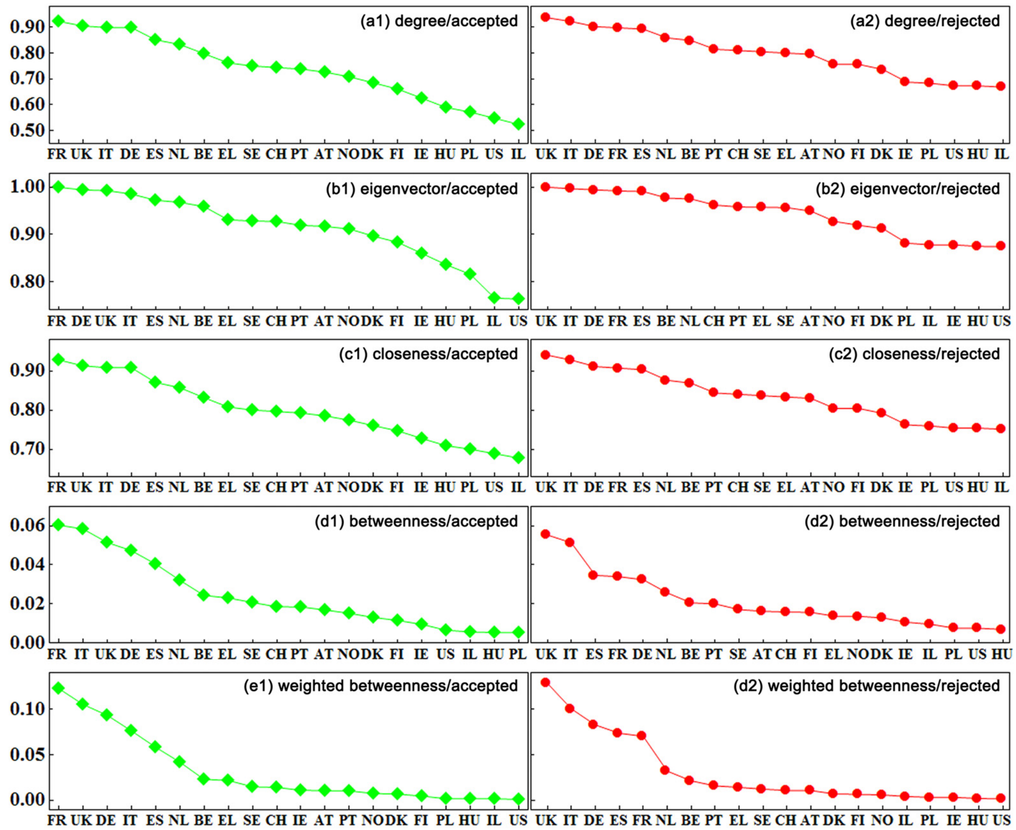

Fig. 4. Comparative ranking of the first 20 countries with the most of signed contracts, according to 5 network centrality indices: (a) degree, (b) eigenvector, (c) closeness, (d) betweenness and (e) weighted betweenness centrality for the networks of participants (in country scale) involved in the FP7 accepted (left column) and rejected proposals (right column).

of the betweenness centrality index, which we will call weighted betweenness and the Minimum Spanning Tree.

Betweenness centrality is a network property based on the notion of shortest path. In reference [21], two ways were proposed for adjusting betweenness centrality to incorporate the weights of the links and thus making it a weighted betweenness. The first suggests to use the weights as a measure of path length. In our case, larger edge weights imply stronger ties between two countries, therefore, its inverse can be perceived as shorter distances between them. One way to incorporate this distance to the weighted betweenness index would be to use the sum of these rescaled values to measure the length of a path. The second alternative considers the weights of the edges as a way to assess the relative importance of a link, rather than using the inverse weights as a measure of the length of the link. The logic behind this can be better understood when we consider the case of a network with edge weights of integer values. In such a network each link is broken down to multiple links, as many as the integer value of its weight. Then, the number of paths connecting any couple of nodes is given by the product of the weights of the consecutive links laying between them. By using the product of the edge weights to count the shortest paths between the nodes of the network, we calculate a weighted version of the betweenness centrality. The exact algorithm for both these alternatives can be found in reference [21]. Of the two alternatives we implement the second. Figure 4 shows the ranking of the twenty countries with the most signed FP7 contracts, according to the five centrality indices we calculated. In order to compare the networks of accepted and rejected proposals, we normalized each centrality. Five countries, France (FR), United Kingdom (UK), Germany (DE), Italy (IT) and Spain (ES) are singled out as the ones with the highest centrality scores and therefore as being the most significant nodes, in both cases. Their relative importance, however, is not the same. For example, node France (FR) ranks first in every centrality in the accepted proposals case, whereas node United Kingdom (UK) does so, in all cases of the rejected proposals. Needless to say, this does neither imply that FR has the most accepted proposals, nor that UK has the most rejected ones.

To assess the relative difference in the importance of each country node in the two networks, we calculated the relative difference

$$\frac{C_{i,acc} - C_{i,rej}}{C_{i,rej}} \tag{3}$$

of every country, where C stands for centrality index, i runs from 1 to 5, to account for each of the five different centralities, acc stands for accepted and rej stands for rejected. This measure can serve as a way to assess the relative difference in the performance of each country

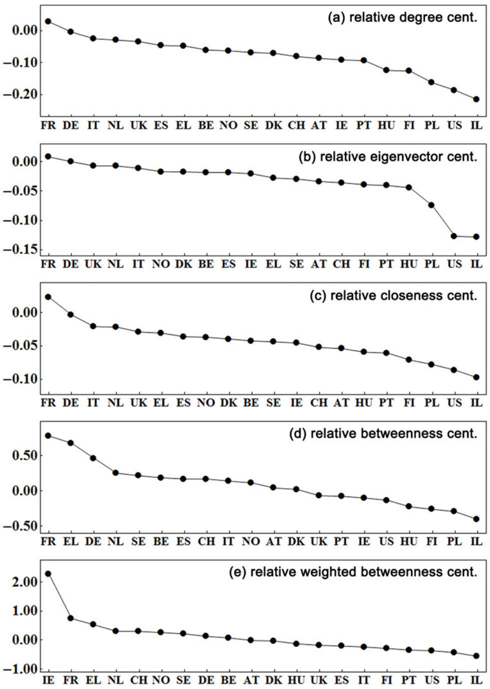

Fig. 5. Comparative ranking of the first 20 countries with the most signed contracts, according to 5 network centrality indices relative difference: (a) degree, (b) eigenvector centrality, (c) closeness, (d) betweenness and (e) weighted betweenness, between the networks formed by the accepted and that of the rejected FP7 proposals.

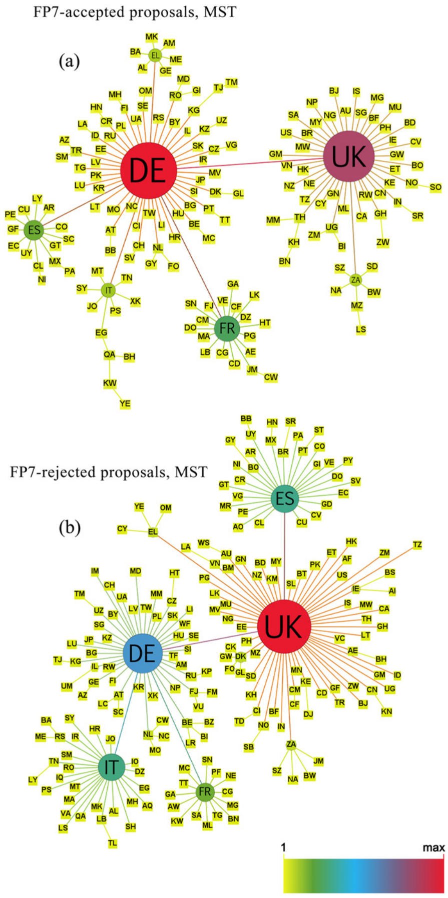

Fig. 6. Minimum Spanning Tree of (a) the network of countries involved in the FP7 accepted proposals and (b) the network of countries involved in the FP7 rejected proposals. The colour (online) of the circles representing the countries varies with respect to the degree of the node in the MST network, according to the chromatic scale shown at the bottom-right part of the figure.

in the two networks, formed by the accepted and the rejected proposals, respectively. The results are shown in Figure 5. As we can see, node France (FR) ranks first according to the relative difference in all cases except the one of weighted betweenness centrality. In this case, node Ireland (IE) exhibits the highest relative difference, with its centrality index value in the accepted proposals case significantly increased compared to that of the rejected proposals one. Assuming that it is preferable for a country to have higher values of centralities in the accepted proposals network rather than the rejected ones, we conclude that France is the country with the best performance overall.

For the construction of the Minimum Spanning Tree, we follow the procedure described in detail in reference [6]. For its implementation we use Kruskal's algorithm [22]. We use the inverse of the link weights, as a measure of the strength of the ties between the countries. The stronger the tie, the shorter the distance. We sort all these distances in a non-decreasing order, then remove all the edges between the nodes. We re-connect the nodes, one pair at a time, starting with the pair with the smallest distance. The outcome of this procedure must be a tree. Therefore, if, at any step, adding the next-in-line edge to the network, would result in a closed path, the particular edge is discarded and we move on to the next. We continue this way, until the entire series of the edges, sorted in a non-decreasing distance order, is taken into account.

Upon sorting the links of the weighted network in a non-decreasing distance order, there are often blocks of equal-distance links distributed throughout the series. As a result, there are as many MSTs as the number of permutations for these blocks of links. To avoid any kind of bias, these blocks of links are shuffled. We average over a sufficient number of different MSTs and compare this average MST to a number of random MSTs to check for any major differences between them. We conclude that there are no notable variations amongst the various equivalent cases of MSTs.

Figure 6 shows a representative example of a Minimum Spanning Tree of (a) the network of the 169 countries, involved in the FP7 accepted proposals and (b) the network of the 206 countries involved in the rejected FP7 proposals. The size of the circles representing the countries vary according to the degree of the node in the MST network. The higher the degree of the country, in proportion to the degrees of the other countries in the network, the larger the radius of the circle. Our results are in agreement to those of previously published work for the FP6 accepted proposals, country-scale network, revealing the same star-like structures around some particular countries [6]. The five most connected countries are Germany (DE), United Kingdom (UK), France (FR), Spain (ES) and Italy (IT), for the accepted proposals case and United Kingdom (UK), Germany (DE), Spain (ES), Italy (IT) and France (FR) for the rejected proposals one. These five countries are those that were also singled out as being the five most important ones, according to the centrality indices, in both networks. Their role, however, in the two MSTs changes, as in the case of the centralities.

4 Discussion and conclusions

We have performed a network analysis of the European Commission projects of FP7, the seven-year period (2007–2013) of supported research, which was recently concluded. We used data provided by the commission, in anonymised form. We separated the data into two sets of proposals, one of the accepted and funded projects and one of the rejected ones and we performed network analysis on the networks that stem from these two sets. In both networks the largest connected components constitutes the majority of the institutes involved in the FP7 proposals. The degree probability distribution functions of the networks of the participants in the accepted and rejected FP7 proposals are indicative of scale-free behaviour. Using the maximum likelihood method we determined that the truncated power-law distribution is the most probable fit for the derived distributions. In both distributions there are three characteristic regions: a region of increasing intensity at small k values ($1 < k < 10$), a linear region at intermediate k values ($10 < k < 1000$) and finally a noisy region for $k > 1000$. We find that proposals with participants having degree k in the first region are liable to be rejected. These proposals are mostly submitted by consortia which are made up of small-scale institutes, which cannot readily satisfy the requirements of a successful proposal on their own. Such consortia could benefit from the addition of one or more larger-scale institutes. In the second region, we see no notable difference, meaning that the value of k does not affect the success of a proposal. In the third region, the data are very scarce and no clear trend can be deduced. By looking at the assortativity coefficient by participant type, we found both networks to be assortative, meaning that there is a trend for academic institutions and for companies to collaborate with partners of the same type. Of the two networks, the one formed by the rejected proposals is more assortative by participant type, an indication that proposals that feature collaborations between academic and industrial participants are favoured. We grouped the participants of the two networks into three geographical scales, namely, the participant's city, region and country and thus constructed six new networks. By evaluating the degree assortativity coefficient of the two networks, we looked at the probability that two nodes with similar degree will be linked and deduced that both FP7 networks are disassortative by degree, in all four scales. This means that in both networks the high-degree nodes preferably link to low-degree ones, creating star-like features in the network. This kind of topology may be attributed to certain circumstances regarding the formation of the collaborations that make up the networks. It also distinguishes them from other social networks which are assortative and present with a dense core consisting of high-degree nodes surrounded from lower-degree nodes. The fact that the FP7 networks exhibit this structural difference conveys two further implications. First, both networks are more susceptible to the removal of high-degree nodes than similar networks with assortative mixing would be. This is evidence supporting that large institutes play a crucial role in maintaining the network's connectivity. Second, the topology of both networks is more efficient for the spreading of ideas, trends, knowledge, technological advances, etc. than that of similar assortative networks. Both networks become more disassortative as their scale becomes more macroscopic, with the exception of the country scale for the accepted proposals network. The network formed by the rejected proposals is more disassortative than that of the accepted, in all four scales, with the highest difference observed in the country scale. Evaluating the average node weight we found that, in both networks, there is a propensity for the average weight to increase with the degree, in agreement to the results of previous studies. The more macroscopic the scale into which we aggregate is, the more observable this propensity becomes. By determining several centrality indices and constructing the Minimum Spanning Trees of the two networks, we elucidated the relative position of each country. We concluded that, although FR, DE, UK, ES and IT are the key players in both networks, their respective influence is not the same. Finally, we calculated the relative difference of the centrality indices, as a way to assess the relative performance of the countries in the two networks. According to this measure, France is recognized as the country with the best overall performance.

We thank the European Commission, Office of Statistics, for providing the data used in this paper. This work was supported by the European Commission FET Project MULTIPLEX No. 317532. We thank Michael Kanetides and Michael Maragakis for their valuable comments.

References

1. L. van Dyck, EMBO Rep. **3**, 1110 (2002)
2. H. Wigzell, Science **295**, 443 (2002)
3. Y. Caloghirou, A. Tsakanikas, N. Vonortas, J. Technol. Transf. **26**, 153 (2001)
4. A. Constantelou, A. Tsakanikas, Y. Caloghirou, Int. J. Technology Management **27**, 773 (2004)
5. J.A. Almendral, J.G. Oliveira, L. Lopez, M.A.F. Sanjuan, J.F.F. Mendes, New J. Phys. **9**, 183 (2007)
6. A. Garas, P. Argyrakis, Europhys. Lett. **84**, 68005 (2008)
7. T. Scherngell, M.J. Barber, Ann. Regional Sci. **46**, 247 (2011)
8. A. Clauset, C.R. Shalizi, M.E.J. Newman, SIAM Rev. **51**, 661 (2009)
9. J. Alstott, E.T. Bullmore, D. Plenz, PLoS ONE **9**, e85777 (2014)

10. M. Newman, Phys. Rev. E **67**, 026126 (2003)
11. B. Efron, SIAM Rev. **21**, 460 (1979)
12. M. Newman, Phys. Rev. Lett. **89**, 208701 (2002)
13. M. Newman, *Networks: An Introduction* (Oxford University Press, Inc., New York, 2010)
14. D. Koschutzki, K. Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl, O. Zlotowski, in *Network analysis: Methodological Foundations*, Lecture Notes in Computer Science, edited by U. Brandes, T. Erlebach (Springer-Verlag, Berlin, 2005), Vol. 3418, pp. 16–61
15. P. Bonacich, J. Math. Sociol. **2**, 113 (1972)
16. P. Bonacich, Am. J. Sociol. **92**, 1170 (1987)
17. G. Sabidussi, Physiometrika **31**, 581 (1966)
18. L.C. Freeman, Sociometry **40**, 35 (1977)
19. J.M. Anthonisse, The rush in a directed graph. Technical report, Stichting Mathemastisch Centrum. Mathematische Besliskunde, BN 9/71, Amsterdam (1971), http://oai.cwi.nl/oai/asset/9791/9791A.pdf
20. U. Brandes, J. Math. Soc. **25**, 163 (2001)
21. U. Brandes, Soc. Networks **30**, 136 (2008)
22. T. Cormen, C. Leiserson, R. Rivest, C. Stein, *Introduction to Algorithms*, 3rd edn. (MIT Press, Cambridge, 2009), http://mitpress.mit.edu/books/introduction-algorithms