Scientometrics (2016) 108:41–56
DOI 10.1007/s11192-016-1952-z

# Evolving cohesion metrics of a research network on rare diseases: a longitudinal study over 14 years

**Carlos B. Amat[1] · François Perruchas[1]**

Received: 25 August 2015 / Published online: 27 April 2016


**Abstract** Research collaboration is necessary, rewarding, and beneficial. Cohesion between team members is related to their collective efficiency. To assess collaboration processes and their eventual outcomes, agencies need innovative methods—and social network approaches are emerging as a useful analytical tool. We identified the research output and citation data of a network of 61 research groups formally engaged in publishing rare disease research between 2000 and 2013. We drew the collaboration networks for each year and computed the global and local measures throughout the period. Although global network measures remained steady over the whole period, the local and subgroup metrics revealed a growing cohesion between the teams. Transitivity and density showed little or no variation throughout the period. In contrast the following points indicated an evolution towards greater network cohesion: the emergence of a giant component (which grew from just 30 % to reach 85 % of groups); the decreasing number of communities (following a tripling in the average number of members); the growing number of fully connected subgroups; and increasing average strength. Moreover, assortativity measures reveal that, after an initial period where subject affinity and a common geographical location played some role in favouring the connection between groups, the collaboration was driven in the final stages by other factors and complementarities. The Spanish research network on rare diseases has evolved towards a growing cohesion—as revealed by local and subgroup metrics following social network analysis.

**Keywords** Co-authorship networks · Social networks analysis · Longitudinal follow up · Rare diseases · Networked research

✉ Carlos B. Amat
cbenito@ingenio.upv.es

[1] Institute of Innovation and Knowledge Management (Ingenio), CSIC and Universidad Politécnica de Valencia, Ciudad Politécnica de la Innovación, 8E Building, Camino de Vera, s/n, 46022 Valencia, Spain

## Introduction and background

Scientific collaboration has been defined as "interaction (…) among two or more scientists that facilitates the sharing of meaning and completion of tasks with respect to a mutually shared, superordinate goal" (Sonnenwald 2007). From a purely scientific perspective, collaboration is necessary to cope with the increasingly challenging, ambitious, and demanding objectives of many research initiatives (in terms of human knowledge as well as material resources). Evidence for growing collaboration comes not only from quantitative data on the rising number of individual contributors per paper (Wuchty et al. 2007) but also from the increasing presence of collective authors. The ATLAS Collaboration in Particle Physics is an extreme example (Cho 2011). In the biomedical research fields, the MONICA Project (cardiovascular health) of the WHO, or the EPSILON study on schizophrenia are good examples of the collective involvement of different countries, centers, and teams in pursuing common research problems. Joint research gives competitive advantage to research partners in the form of more frequent citations for co-authored papers (Larivière et al. 2015) and therefore greater funding opportunities.

Collaborative research also has beneficial socioeconomic outcomes (OECD 2010, p. 27) and so national and supranational public entities are either encouraging research partnership in the implementation of specific programs or introducing criteria to foster collaborative approaches in funding applications. The cross-national COST actions that "fund pan-European, bottom-up networks of scientists and researchers across all science and technology fields" (http://www.cost.eu/about_cost/how_cost_works), or the Clinical and Translational Science Awards program of the US National Center for Advancing Translational Sciences (https://www.ctsacentral.org/) are good examples of the former; while the principles for funding multi-institutional collaboration in innovation and research published by the UK Research Council (http://www.rcuk.ac.uk/funding/principles/) is a good example of the latter.

Fostering agencies need instruments to assess the efficiency of research collaborations. In fact, a recent report on so-called "team science" recommends that researchers "partner with team science leaders to evaluate and improve analytical methods and tools for team assembly" (Cooke and Hilton 2015) and network analysis has emerged both as an adequate framework and as a convenient analytical tool to understand these processes and evaluate their outcomes (Hunt et al. 2012; Bian et al. 2014).

### Social network analysis of collaboration

A network is a representation of a system. It consists of vertices that represent the entities of the system. Pairs of vertices are joined by edges that represent a particular kind of interconnection between these entities (Estrada 2011). Citation relationships between documents are considered the source of a type of knowledge network, namely, citation networks (Newman 2003b); while the co-authorship of articles in learned journals (Newman 2001a, b) along with the networks constructed from collaborative research grant applications (Bian et al. 2014) have been studied as examples of social networks. Throughout this paper, we use the term network instead of net or graph; likewise, we use teams to refer to the research groups that are connected or linked by edges.

In co-authorship networks, individuals or entities (teams, institutions, etc.) are connected if they have co-authored one or more papers. Such a simple relationship and the resulting systems have attracted a considerable number of works from Scientometrics, Social Network Analysis (SNA), and other research fields. Co-authorship networks have been approached either in a longitudinal or a cross-sectional manner and analyses have been performed at the collective or elemental level. We will next review some examples of these approaches and their utility for our purposes.

Mark Newman can be credited for undertaking the first large-scale analysis of scientific co-authorship (Newman 2001a, b). Although his main aim was to obtain a reliable social network based on the assumption that joint authorship reflects genuine professional interaction between scientists, the metrics he used for characterizing the networks (we will review these in the appropriate section) have remained as a model for analyzing co-authorship networks at the collective level. His study, however, is cross sectional, and reduced to a 5-year period with accumulated figures. María Bordons and colleagues, in another example of transversal research, relate the research performance and the network position of individual researchers in pharmacology, nanoscience, and statistics (Bordons et al. 2015). There are two other evaluative papers focused on the participation of research groups in the Clinical and Translational Science Awards: the already mentioned article on the University of Arkansas for Medical Sciences (Bian et al. 2014) uses SNA for "evaluating the impact of resource allocation to different programs"; and an article on the Indiana Clinical and Translational Sciences Institute (Hunt et al. 2012) "derives a single consented ranking of important or influential nodes in a collaboration network".

Xiaoming Liu and co-workers analyzed the structure of collaboration within the Digital Libraries research community and provided quantitative metrics for the concepts of status and influence of individual authors (Liu et al. 2005). They added to the individual prestige metrics some general measures of the whole structure of the network; however, their analysis remains static. A good example of a longitudinal study at the elemental level could be the analysis of "social inertia" in which Ramasco and Morris (2006) follow 14 research collaboration networks (and another derived from the Internet Movie Database) during an indefinite time period.

Shortly after the seminal work by Newman, several other statistical physicists studied the evolution of social networks using co-authorship networks as examples (Barabási et al. 2002), although to our knowledge Newman's paper was the first example of dynamic analysis of co-authorship, their main interest seemed to be the large scale modeling of complex evolving networks. In contrast, the work by Börner et al. (2005) appears to be the first that combines the positional metrics of individuals in the network with several "success" indicators of the papers they contribute and, more importantly, a longitudinal follow-up of the characteristics of the whole network. In the same line, Luis Bettencourt and co-authors follow the emergence and development of a series of research specialties tracking several co-authorship network metrics as the corresponding fields evolve (Bettencourt et al. 2009). More recently, Ghosh and collaborators followed several structural measures of evolving co-authorship networks however; their treatment of network cohesion is quite superficial (Ghosh et al. 2014).

Several sets of metrics have been applied, in the above mentioned and other works, to characterize either the behavior of the whole co-authorship network and its evolution, or the role of individual nodes whose topology reveals an outstanding position, influence, or importance in the observed research field. These network and node metrics have been combined with "efficiency" estimates, usually in the form of research outputs and impact indicators that attempt to confirm the benefits of collaboration. However, before we move on with a detailed review of these measures and state our objectives, it is worthwhile describing the context in which our analysis takes place.

## CIBERER as our case of study

Biomedical Research Networking Centers (CIBER after the Spanish acronym) is a Spanish public initiative to support single-topic research on specific broadly-defined disease or health problems. Following an initial call in 2006, nine monographic centers were established on Neurodegenerative Diseases, Hepatic, and digestive diseases, Public health and epidemiology, Bio-engineering, biomaterials and nanomedicine, Diabetes and associated metabolic disorders, Physiopathology of obesity and nutrition, Mental health, Respiratory diseases and Rare diseases. This last mentioned center, CIBERER for short, is the object of our work.

The word "center" might be misleading, as every consortium is made up of a number of research teams from various parent organizations. From its beginning in 2007, 61 research teams joined CIBERER and, after one was removed in 2009, the consortium included 60 teams, encompassing 700 people in 2013. The last annual report details an annual budget (in 2013) of 6.7 million euros coming from the hosting public institutions and research program funding plus 1.5 million euros from private funding, which comes to a total of 8.2 (or almost USD 9.4 million at the current exchange rate).

The starting point for the CIBER program is "the need to boost research excellence through the implementation of stable structures of research collaboration" (Ministerio de Sanidad y Consumo 2006). This main goal adds to a specific feature of rare diseases that is implicit in its name: this is a very large group of diseases (over 7000 following international criteria) with a low prevalence in the population. Most are oncological, neurological, or metabolic disorders—usually with genetic origins. So it should come as no surprise that the need for networked research on rare diseases had already been stressed (Aymé and Schmidtke 2007). Collaboration in rare diseases research comes, then, as a twofold necessity. Firstly, as the number of patients is small, it is necessary to cooperate to avoid the fragmentation of research and gain shared knowledge. Secondly, it is necessary to cooperate to associate the clinical features and, eventually, identify new genes or associate gene expression and mutation consequences to the clinical features of a given disorder. Hence, the need for organizational plus instrumental collaboration adds to the call for "excellence" through the foundation and development of institutional collaboration. CIBERER encourages collaboration between their own groups through intramural programs and cooperative actions. In 2012 they were launched nine such internal programs and by 2014 the number rose to 13.

How could we measure both the degree of collaboration and the effects eventually resulting from joint research efforts? The next paragraph introduces some concepts from network analysis and metrics that may help represent collaboration relationships and measure their strength and evolution. For the sake of comprehension, the concepts used are loosely-defined.

Network performance is related to network cohesion. Several reviews and meta-analyses on cohesion-performance relationship have found a positive correlation, although this correlation is moderate and highly dependent on intra-group processes (Chiocchio and Essiembre 2009). We approach the concept of group cohesion in a loose and pragmatic way, as the group members' inclination to forge, maintain, and even reinforce social ties in order to achieve common goals and mutual benefits while group integrity is preserved (Casey-Campbell and Martens 2009).

A network is a network insofar as it connects its members. The main purpose of this paper is to follow the collaboration relationships among the research teams of a formal research consortium on rare diseases and examine if the collective evolution leads to greater cohesion. This will be accomplished by first determining the research output of the groups and identifying the journal publications that two or more groups co-author. Secondly, we will build the collaboration networks and apply network metrics to observe the collective evolution of the groups and their interrelationships. We are particularly interested in those measures that best reflect the above definition of network cohesion. Thirdly, we will offer some concluding remarks that could orientate future research on scientific co-authorship networks.

## Basic concepts, dataset and analysis

Collaboration and co-authorship are far from equivalent concepts (Laudel 2002) although Scientometric practice has sanctioned the use of the latter as a proxy for the former. Here, we refer to the relationship between two investigators who jointly appear on the byline of a research paper as co-authors. When aggregated to the institutional or even higher level, as is the case with this article, we refer to collaboration as, say, two universities collaborating rather than co-authors collaborating in one or more papers. Our network takes teams as the entities of analysis (nodes or vertices in the SNA terminology) and establishes connections (edges) between pairs of teams if their members have co-authored one or more papers. Co-authorship edges show some intensity according to the number of co-authored papers between teams. The total number of papers co-authored by a team is its strength or weighted degree. As an example, let us look at the small network on the left side of Fig. 1. The vertex labelled ‘b’ connects with three other vertices (a, c and d) in the way that vertex d does; but b has more strength despite having the same degree because the weight of the edge b–a is two. Both have degree 3 but b has strength 4. Although the strength distribution gives some idea about the relationship among entities, a better understanding of the whole network can be drawn from group level metrics—especially if these metrics are followed along the same period.

On the right side, Fig. 1 depicts the same network but at a more advanced stage and several changes are quite evident. The most obvious is that the two components on the left (vertices h and i) have coalesced into one component that connects all the vertices. Another

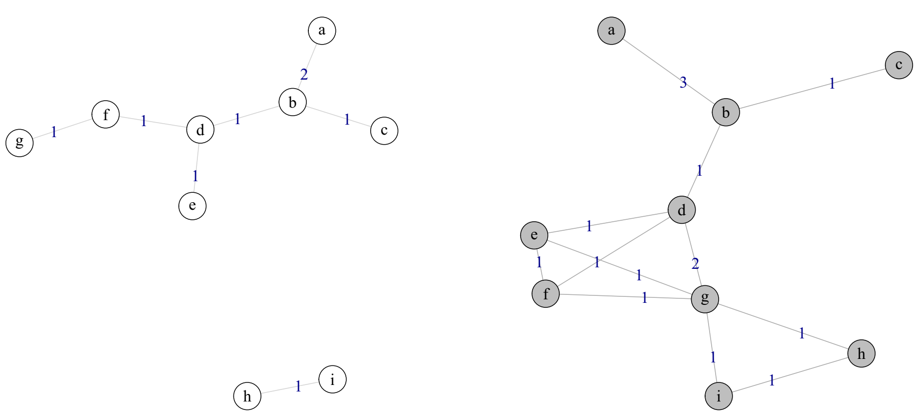


**Fig. 1** Successive stages of an ideal network evolving (from *left* to *right*) towards a greater cohesion

less obvious change is the appearance of triads: sets of three fully connected vertices from a previous situation where they were connected only in couples. This, for example, is the case of vertices d–e–f and f–g–d. Finally, it is obvious that a complete subnetwork has been developed among the four vertices d–g, with each connected to the others.

We studied networks established among CIBERER research teams that co-authored papers. Although CIBERER was formally established late in 2006 and started activities in 2007, we extended the scope of our study from 2000 because it was important to identify some "invisible college" effect, that is, some previous interactions already in place among the member groups. In that case, the foundation of collaboration could be hardly attributable to the formal network activities. We follow the CIBERER call in defining a team or research group as a set of researchers around a principal investigator who collaborate in the study of a homogeneous topic (Ministerio 2006) within an organizational system, like a hospital, a research facility and the like. The CIBERER website (www.ciberer.es) provides a list of all involved teams, every P.I. and a full list of members.

Using author and affiliation data, we identified and retrieved from the Web of Science (WoS) 4710 journal publications contributed by CIBERER teams during the period 2000–2013. As the resulting bibliographic lists used to be incorporated to the web site of CIBERER (www.ciberer.es) on a periodical basis, every group had the opportunity to check for the accuracy and completeness of their own bibliographies and to ask eventually for amendments.

In 987 of these papers we identified two or more co-authoring CIBERER teams, meaning that the authors of these papers were affiliated to two or more groups. As we were only interested in the collaboration inside this consortium, we disregarded any other co-authoring data in our analysis.

After processing the bibliographic dataset, we obtained two tabular files for every year in the period 2000–2013. The first contained the groups who published at least one paper in that year. Along with their identification, this vertex file listed every team along with their specific research areas, the geographical location of their host institutions, their clinical or basic orientation, and the number of papers contributed. Only teams publishing at least one paper in a given year were included in the corresponding set. The second file was a plain list of the pairs identified in the papers for the same year. Those familiar with network analysis would recognize the two components of the standard file format used by most applications. We used the R package iGraph (Kolaczyk and Csardi 2014) to run the analytical routines on the data, which included some very basic distributions, such as the degree distribution, as well as network cohesion metrics as they evolved in the time frame.

# Results

## Research output

Before discussing the network analysis, it seems convenient to give some data on the research output of CIBERER and its groups. We identified 4710 journal publications contributed by the 61 research teams between 2000 and 2013. The Web of Science subject categories most frequently attributed to the papers were by far Genetics and Heredity and Biochemistry and Molecular Biology, then came Endocrinology and Metabolism, Neurosciences, Clinical Neurology, and Oncology. This distribution gives some idea of the cross-disciplinary composition of the consortium.

Figure 2 summarizes the research output of the groups in terms of published papers per team during the period. The average number of papers per team and year grew from 4.46 to 10.1 over the period with standard deviations of 2.97 and 9.46, respectively; these figures are consistent with the evolution of the median of the distribution: in 2000 half of the groups published four papers at least—while in 2013 they doubled their output and published eight papers. Dispersion grew progressively during the period; the initial interquartile range was four, while leaving aside the years 2008 and 2009, the range reached and even exceeded seven by the final year of the series. Another imbalance is revealed by the progressive appearance of outliers who, in the two latter years of the period, reached a maximum output of more than 50 papers a year. It seems that, regarding their publications output, setting up the consortium caused a growing inequality among the groups.

On the other hand, Fig. 3 presents the dramatic increase in the number of foreign institutions (here expressed on a logarithmic scale) contributing to CIBERER papers in the last years of the period. For 2000–2002, half of CIBERER papers just showed one foreign collaborating institution (log10 = 0). The median of the distribution rose to 0.3 (meaning two foreign institutions) between 2003 and 2011 while, at the same time, the outliers proliferated. In the last 2 years of the series, half of the papers had three international contributing institutions or more and the great number of cases which outlay have displaced the arithmetic mean above four foreign institutions per paper. The extreme case in 2012 corresponds to a consensus paper (PMID 22966490) with more than 1150 contributing institutions.

## Cohesion analysis

Many co-authorship network analysis begin by looking over the number of connections incident to the nodes. While this approach is reasonable in cross-sectional studies focusing on individual elements of the net, centrality metrics looking to identify outstanding nodes are of little use in studying the collective behavior of evolving networks. It makes much

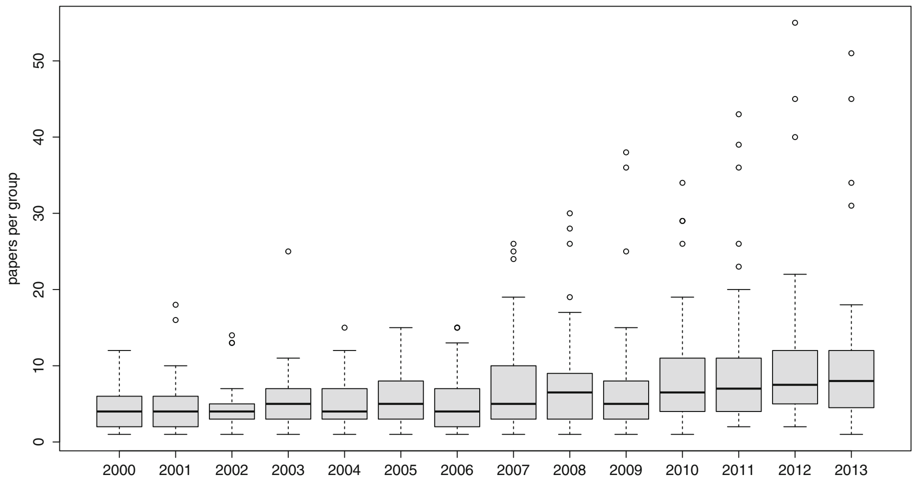


**Fig. 2** The research output of CIBERER in terms of the distribution of papers per group for each year of the period. Full data is available from the authors

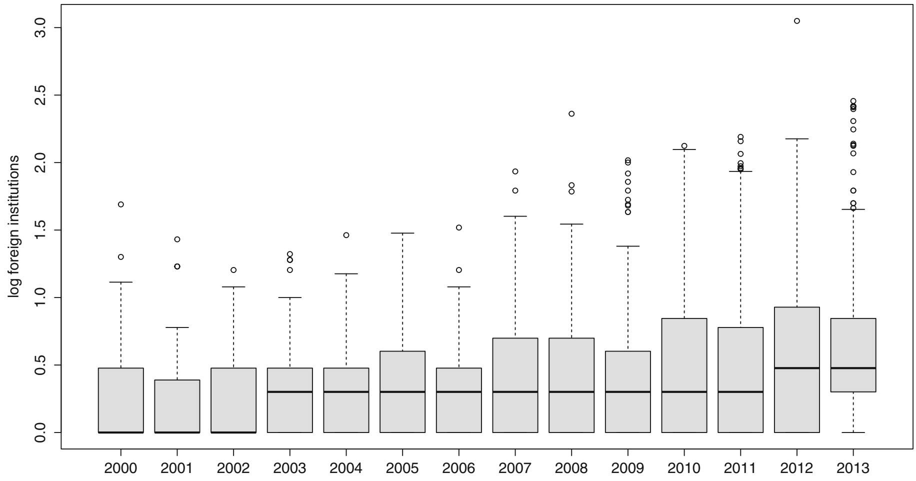


**Fig. 3** The frequency distribution of the number of foreign institutions per paper on a logarithmic scale shows its dramatic increase, particularly in the two last years. As always, full data is available upon request

more sense to pay attention to cohesion metrics because those values reflect the evolution of CIBERER groups towards either coalescence or disaggregation. Table 1 gives a summary of the activity of the teams and also of their relationships throughout the period. Three series of values are provided for every year. The first one includes global measures, the second one relates to the local features of the evolving network; finally, a set of organizing patterns is detected in the shape adopted by some topological features. The first row in Table 1 contains the number of active teams for each year, meaning teams that published at least a one paper in that year. Since the formal constitution of the consortium in 2007, almost the 97 % of the groups are active. The share was much lower in 2000, with less than 86 %, and rose to a 93.4 % in 2006. By 2007, the year the shared activity took off, all the groups where active in publishing research papers and this feature has remained stable except for 2011 and 2013.

## Global measures

Although most studies use the number of co-authoring authors or groups as the main indicator of the intensity of collaboration, it is the number of connections that best reveals how a network evolves. In fact, the average number of groups per paper was 1.13 in 2000, 1.15 in 2007, and 1.18 in 2013 (confidence interval 1.13–1.21 for this last year) while the number of connections almost tripled for the number of active groups which grew 10 % between the extreme years. Indeed, the average number of co-authored papers, denoted by the average strength row, grew from less than three in 2007 to more than four in 2013 after peaking in 2012 (while the period 2000–2006 presents more modest figures). Figure 4 shows the evolution of the distribution of the tie strength among the groups.

The proportion of connections established to the total number of possible connections is called the network density and, in our case, this density increased 2.23 times during the period—however, it multiplied by 1.59 between 2000 and 2006 and by a mere 1.63 in the formal period. A commonly used quantity in network literature is the ratio of the total number of triples that form triangles [as with the g–h–i nodes in Fig. 1(right) over the total

**Table 1** Global, local and community metrics for each annual co-authorship network

| | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| No. of teams | 52 | 58 | 57 | 57 | 59 | 58 | 57 | 61 | 61 | 61 | 60 | 59 | 60 | 59 |
| No. of connections | 31 | 34 | 30 | 48 | 46 | 47 | 58 | 54 | 77 | 57 | 87 | 89 | 87 | 86 |
| Density | 0.022 | 0.021 | 0.019 | 0.028 | 0.027 | 0.028 | 0.035 | 0.030 | 0.041 | 0.031 | 0.049 | 0.050 | 0.049 | 0.049 |
| Transitivity | 0.296 | 0.452 | 0.207 | 0.432 | 0.429 | 0.344 | 0.435 | 0.310 | 0.479 | 0.357 | 0.411 | 0.213 | 0.266 | 0.221 |
| Components | 3 | 4 | 5 | 6 | 4 | 5 | 4 | 3 | 3 | 4 | 3 | 3 | 3 | 2 |
| Max component | 20 | 13 | 13 | 19 | 19 | 23 | 31 | 32 | 46 | 36 | 41 | 49 | 51 | 50 |
| Max component density | 0.137 | 0.179 | 0.154 | 0.164 | 0.164 | 0.103 | 0.105 | 0.091 | 0.073 | 0.086 | 0.104 | 0.075 | 0.067 | 0.070 |
| Cliques | | | | | | | | | | | | | | |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| 4 | 1 | 2 | 0 | 0 | 2 | 1 | 0 | 3 | 2 | 0 | 5 | 0 | 1 | 0 |
| 3 | 3 | 3 | 2 | 6 | 7 | 7 | 7 | 5 | 8 | 10 | 9 | 28 | 15 | 17 |
| 2 | 18 | 15 | 24 | 20 | 16 | 22 | 19 | 29 | 24 | 23 | 22 | 26 | 32 | 32 |
| Maximum distance | 9 | 7 | 10 | 10 | 10 | 11 | 13 | 13 | 11 | 11 | 11 | 11 | 14 | 13 |
| Mean distance | 3.179 | 2.482 | 2.800 | 2.973 | 3.163 | 2.276 | 3.467 | 3.579 | 3.854 | 3.697 | 3.139 | 3.142 | 3.373 | 3.408 |
| Average strength | 1.481 | 1.690 | 1.509 | 2.305 | 2.339 | 3.220 | 2.931 | 2.721 | 3.516 | 2.721 | 4.867 | 4.667 | 5.100 | 4.067 |
| Communities | 32 | 34 | 31 | 27 | 29 | 23 | 24 | 28 | 20 | 29 | 22 | 17 | 15 | 18 |
| Average no. of members | 1.687 | 1.705 | 1.838 | 2.185 | 2.034 | 2.521 | 2.416 | 2.178 | 3.100 | 2.103 | 2.727 | 3.529 | 4.000 | 3.333 |

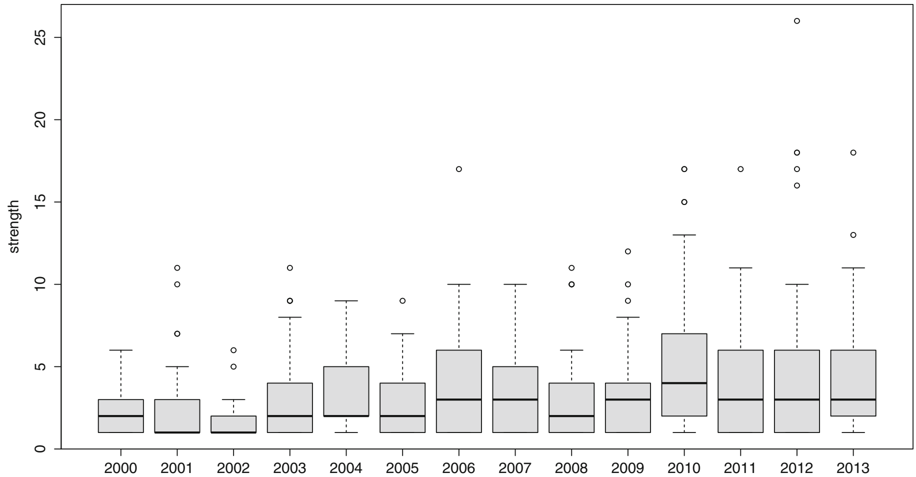


**Fig. 4** The evolution of vertex strength (weighted degree) of the groups along the period. Full distribution is available from the authors

number of connected triples, as with a–b–c]. This is termed transitivity or clustering coefficient, and the values fluctuate during the period with no clear tendency.

## Local measures and subgraphs

A network is connected if there is a path (a sequence of nodes and ties) between each pair of nodes. Figure 1(right) is a connected network while the network on the left has two components. It is obvious that the more components, the less cohesive is a network. The giant component is the subnetwork with the larger number of connected groups. The corresponding values for these two metrics are shown in Table 1 and although the number of components has remained practically unchanged since 2007, it is the size of the giant component that reveals how cohesive the CIBERER network has become. In 2007, the first year of collective activity, the giant component involved 52.46 % of the active groups; in 2013 this share rose to 83.33 %. In the early years, only 2006 showed a share above 50 %. The emergence of a giant component is not the only indicator at the local level of a rise in connectivity and, hence, network cohesion. The network is complete when every node is connected to every other node. A complete subgraph is called a clique, as is the case with subgraphs d–e–f–g and g–h–i in Fig. 1(right). Cliques in Table 1 are classified by the number of interconnected teams—from two to six. It becomes apparent that the formal period is populated by cliques of three to five groups, with cliques of order six in 2008 and 2010.

## Communities and mixing

In addition to the local topology of a network and its subgraphs, there are some results related to the patterns of interrelation among the nodes, the so-called communities. A community is formed by an internally connected set of nodes for which the internal density is significantly larger than the external density (Estrada 2011). The last two rows of Table 1 contain data on the evolving community structure of the CIBERER network. The

**Table 2** Assortativity coefficients with regard to location, subject field and profile of the CIBERER research groups

| | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject field | | | | | | | | | | | | | | |
| Endocrine medicine | NaN | −0.015 | 0.649 | −0.011 | −0.011 | −0.011 | 0.658 | −0.029 | −0.027 | 1.000 | 0.382 | 0.661 | 0.310 | −0.006 |
| Mitochondrial medicine | −0.071 | 0.244 | 0.346 | 0.400 | 0.158 | 0.365 | 0.308 | 0.388 | 0.396 | 0.475 | 0.297 | 0.352 | 0.232 | 0.066 |
| Genetic medicine | 0.774 | −0.578 | 0.683 | 0.289 | −0.436 | 0.736 | −0.357 | 0.112 | 0.279 | 0.156 | 0.215 | 0.203 | 0.069 | 0.017 |
| Hereditary metabolic medicine | 0.262 | 0.230 | 0.280 | −0.050 | 0.198 | 0.274 | 0.235 | 0.509 | −0.032 | 0.261 | 0.372 | 0.040 | 0.142 | 0.101 |
| Neurosensory pathology | 0.139 | 0.150 | 0.085 | −0.067 | 0.138 | 0.105 | 0.342 | 0.232 | 0.188 | −0.075 | 0.206 | 0.190 | 0.036 | 0.051 |
| Hereditary cancer | 0.650 | 0.651 | 0.345 | 0.330 | 0.326 | 0.329 | 0.240 | 0.190 | 0.410 | 0.219 | 0.343 | 0.170 | 0.504 | 0.270 |
| Pediatric and developmental medicine | 0.466 | −0.079 | 0.040 | 0.076 | −0.095 | 0.386 | 0.071 | 0.250 | 0.303 | 0.211 | 0.112 | 0.422 | 0.322 | 0.078 |
| Geographical location | | | | | | | | | | | | | | |
| Madrid | 0.354 | 0.348 | 0.306 | 0.378 | 0.428 | 0.374 | 0.328 | 0.366 | 0.338 | 0.307 | 0.172 | 0.143 | 0.140 | 0.181 |
| Barcelona | 0.320 | 0.393 | 0.657 | 0.418 | 0.617 | 0.317 | 0.194 | 0.410 | 0.230 | 0.200 | −0.092 | 0.148 | 0.158 | 0.029 |
| Valencia | −0.408 | 0.469 | 0.630 | 0.048 | −0.095 | 0.603 | 0.025 | 0.232 | 0.260 | 0.441 | 0.376 | 0.099 | 0.114 | 0.155 |
| Clinical or basic profile | −0.408 | 0.056 | −0.010 | 0.153 | 0.141 | 0.105 | 0.125 | 0.175 | 0.015 | −0.019 | −0.012 | −0.040 | −0.044 | −0.031 |
| Assortativity (degree) | 0.035 | 0.159 | 0.321 | 0.218 | 0.117 | −0.055 | −0.025 | 0.134 | 0.246 | 0.186 | 0.125 | −0.069 | 0.043 | −0.029 |
| Research output | −0.093 | −0.280 | 0.193 | 0.109 | −0.018 | −0.024 | −0.126 | −0.008 | 0.120 | 0.200 | 0.093 | 0.058 | 0.109 | 0.066 |

most obvious finding is the sharp decline in the number of communities since the consortium's start and an increase in the average number of members. This result is consistent with the increase in higher order (≥3 members) cliques and the average number of members of each community. To further investigate this logical organization of the net, we examined whether there was a tendency for the groups to connect with alike others. This feature is called homophily (or assortativity) and, like the correlation coefficient, its values range from −1 to 1, depending on the degree of association of nodes with respect to some nominal or scalar variable (Newman 2003a). We have calculated the assortativity coefficient of the networks throughout the whole period with respect to the degree of the nodes and several attributes such as output (number of papers), location, research field, and basic or clinical profile. With regard to this last feature, teams were classified according to their host institutions. So, a group pertaining to a research institution (say Spanish High Research Council) or to an academic one (Autonomous University of Madrid, for instance) was classified as basic groups while those working in hospitals were considered as clinically oriented.

The values appear in Table 2 and show, with respect to the number of collaborations of every group and the number of published papers, values close to zero—meaning a lack of influence on the propensity to collaborate. Most of the research groups were concentrated in the three cities included in the table and the coefficients seem to indicate that geographical proximity plays a significant role in the first period, while diminishes in the most

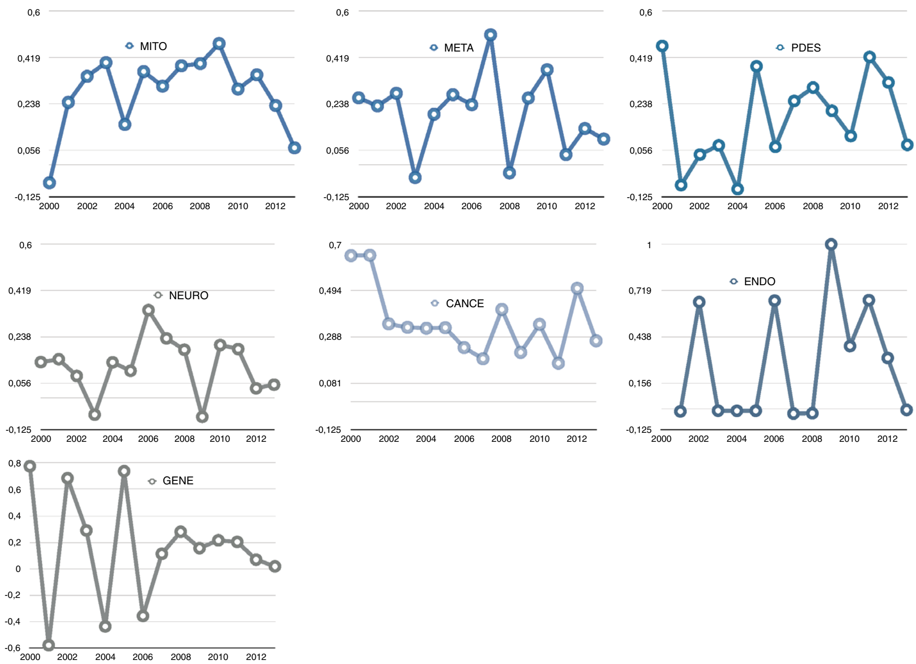


**Fig. 5** The propensity of the teams to collaborate following their subject orientation, measured by the values of the assortativity coefficient in each subject area throughout the period. Clockwise from *top left*, mitochondrial medicine (MITO) hereditary metabolic medicine (META) pediatric and developmental medicine (PDES) neurosensory pathology (NEURO) hereditary cancer (CANCE) endocrine medicine (ENDO) and genetic medicine (GENE)

recent years. A similar decline in the coefficients can be seen with regard to the clinical or basic orientation of the groups but, in this case, negative values might suggest a propensity to collaborate between basic and clinical groups, which fit with the objectives of the consortium. Moreover, the ascription of some groups to the same research line follows several patterns, which are visible in the curves of Fig. 5. The most conspicuous variation appears in the case of hereditary cancer, a highly assortative set of seven groups in 2000 whose thematic relationship seems to have vanished as time went by. This is also the case for the pediatric groups; however, it started from a lower assortativity coefficient based on its thematic profile. It’s noteworthy to observe how in several cases the values peaked around the time (2006–2007) the network was formed and then dropped suggesting an inverse U shape of the distributions. In all but the Endocrinology and Mitochondrial diseases fields, the initial coefficients were higher or much higher than the final ones and it is also apparent a general change from the somehow erratic values of the previous stage to the attrition of the latter.

These results suggest that the teams were well connected before the formal constitution of the network and that they were linked because of their subject affinity and even their geographic location. The values corresponding to the teams based in Valencia, grouped in two main hosting institutions, are consistent with this first stage measures. However, once the network was in place, the teams begin to collaborate on a different basis. This fact along with the lack of connectivity based on the clinical or basic character of the groups is good news for CIBERER, as it seems to suggest a turn towards a collaboration not just

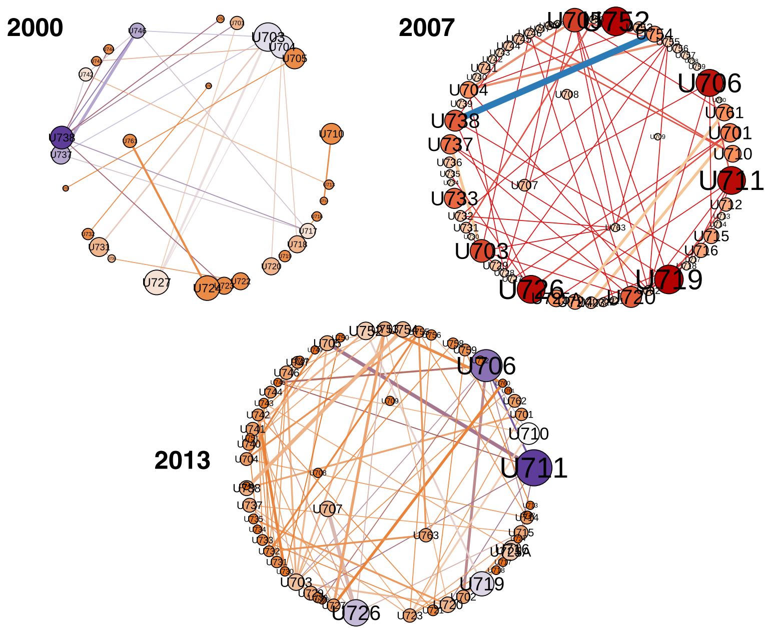


**Fig. 6** Three snapshots of the CIBERER co-authorship network. In these dual circle layouts (drawn with Gephi) nodes represent teams and have been sized following the research output of every group. Isolated (non collaborating) teams have been removed. *Thickness of edges* denotes intensity of collaboration. Nodes have been arranged and can be identified by their labels. *Darker colors* signal more connected nodes. (Color figure online)

based in previous acquaintances but in current common interests and research complementarities.

### First dates

When we observed the year in which two groups first co-authored a paper, we found that more than a half (54.67 %) of these 300 "first acquaintances" took place in 2007 or later, which reveals the influence of the consortium on the behavior of the groups.

The three-part Fig. 6 captures three moments of the evolving collaboration among CIBERER groups. In every dual circle layout, drawn with Gephi (Heymann 2014) the nodes represent the research teams and the edges link those who collaborate in that year. The diameter of the nodes is proportional to their research output and the color gradation signals the intensity of the collaboration as well as the number of different teams one specific group is co-authoring papers with. Nodes are labelled in order to follow changes from year to year. As only the collaborating groups are depicted, it is evident from these layouts the progressive incorporation of the units to the common goal and the growing connectivity from earlier stages.

Finally, he observed collaboration behavior (and measures) could have been affected by a prolonged publication delay of papers, but a quick estimate of the delay between reception and publication dates of the CIBERER papers appeared in 2014 revealed an average delay of 158 days (the maximum value was 379). On the other hand, the acceptance delay of papers published in CIBERER most frequented journals (Journal of Inherited Metabolic Diseases, PLoS One, Forensic Science International…) rounds 100 days, according to some recent data (Himmelstein and Powell 2016) so, it's difficult to think about some displacement of the effects of funding or encouraging collaboration above 1 year.

## Conclusions

The analysis of Spanish research network on rare diseases reveals a growing cohesion. We have detected some level of interaction among the groups before the network was formally founded. However, and despite the apparent steadiness that some global measures indicate, there is a clear tendency towards coalescence as revealed by the progressive reduction of the number of separate components along with the emergence of a giant component. This giant component has been observed in other evolving co-authorship networks but using aggregate data (Kumar 2015) while in the case of CIBERER it is visible in the every year network.

Indeed, the appearance of fully connected subgroups (cliques) of higher order and the reduction of the number of separate communities combined with the increase in the average number of members are additional indicators of greater network cohesion. In the end, a network is made up of connected elements; otherwise it is not a network. These results are consistent with those of Cugmas et al. (2015) who confirm the hypothesis of an increase in average core sizes as networks both in natural and in social sciences evolve. Furthermore, the finding of "chained communities" (Liu and Xia 2015) is also consistent with the groups we have identified, although we can't confirm at the moment their assembly in a giant component eventually leading to a small word structure, as these Chinese authors do.

Although we have intended to deal with the collective behavior of the CIBERER network, we must recognize that its evolution, in terms of research output, has led some teams to an outstanding position while others remain at a rather discreet level. The successful groups seem to have taken more advantage from their participation in the consortium. A longer follow up could lead to a more balanced picture as more teams progress to higher levels, but we will always expect some degree of inequality as it seems to be the rule with these data. Maybe this is a parallel evolution to the threefold structure devised by Liu and Xia (2015) and similar to the "semi-periphery" and "periphery" structures that Cugmas et al. (2015) obtain applying block-modeling to the networks of Slovenian researchers.

The analysis of homophily, revealing a change in the pattern of connections, might be used to unveil links among the nodes of co-authorship networks not due to previous mutual knowledge of the teams but arising from some just discovered complementarities, like instrumental ones. In fact, what seems to be the rule with some other homogeneous networks (Gallivan and Ahuja 2015) isn't necessarily a good indicator regarding the evolution of the net towards greater cohesion.

Finally, let us quote the team science report that says "As teams and groups develop and move through their phases of scientific problem-solving, their interactions will change, and the field must identify how to measure these team processes". Bibliometrics and co-authorship network analyses are explicitly mentioned, along with other qualitative methods such as the techniques whose combination is needed to ascertain how team processes are related to the multiple goals of transdisciplinary team science (Cooke and Hilton 2015, p. 52). Our ultimate goal has been to contribute with our results to this emerging field of research.

**Acknowledgments** The Spanish Ministry of Economics and Competitiveness partially supported this research (Grant Number ECO2014-59381-R).

## References

Aymé, S., & Schmidtke, J. (2007). Networking for rare diseases: A necessity for Europe. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz, 50*(12), 1477–1483. doi:10.1007/s00103-007-0381-9.

Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications, 311*(3–4), 590–614. doi:10.1016/S0378-4371(02)00736-7.

Bettencourt, L. M. A., Kaiser, D. I., & Kaur, J. (2009). Scientific discovery and topological transitions in collaboration networks. *Journal of Informetrics, 3*(3), 210–221. doi:10.1016/j.joi.2009.03.001.

Bian, J., Xie, M., Topaloglu, U., Hudson, T., Eswaran, H., & Hogan, W. (2014). Social network analysis of biomedical research collaboration networks in a CTSA institution. *Journal of Biomedical Informatics, 52*, 130–140. doi:10.1016/j.jbi.2014.01.015.

Bordons, M., Aparicio, J., González-Albo, B., & Díaz-Faes, A. A. (2015). The relationship between the research performance of scientists and their position in co-authorship networks in three fields. *Journal of Informetrics, 9*(1), 135–144. doi:10.1016/j.joi.2014.12.001.

Börner, K., Dall'Asta, L., Ke, W., & Vespignani, A. (2005). Studying the emerging global brain: Analyzing and visualizing the impact of co-authorship teams. *Complexity, 10*(4), 57–67. doi:10.1002/cplx.20078.

Casey-Campbell, M., & Martens, M. L. (2009). Sticking it all together: A critical assessment of the group cohesion–performance literature. *International Journal of Management Reviews, 11*(2), 223–246. doi:10.1111/j.1468-2370.2008.00239.x.

Chiocchio, F., & Essiembre, H. (2009). Cohesion and performance: A meta-analytic review of disparities between project teams, Production teams, and service teams. *Small group research, 40*(4), 382–420. doi:10.1177/1046496409335103.

Cho, A. (2011). Particle physicists' new extreme teams. *Science, 333*(6049), 1564–1567. doi:10.1126/science.333.6049.1564.

Cooke, N. J., & Hilton, M. L. (2015). Enhancing the effectiveness of team science. Washington, D.C.: National Academies Press. Recuperado a partir de http://www.nap.edu/catalog/19007/enhancing-the-effectiveness-of-team-science.

Cugmas, M., Ferligoj, A., & Kronegger, L. (2015). The stability of co-authorship structures. *Scientometrics, 106*(1), 163–186. doi:10.1007/s11192-015-1790-4.

Estrada, E. (2011). *The structure of complex networks: Theory and applications*. Oxford: University Press.

Gallivan, M., & Ahuja, M. (2015). Co-authorship, homophily, and scholarly influence in information systems research. *Journal of the Association for Information Systems, 16*(12), 980.

Ghosh, J., Kshitij, A., & Kadyan, S. (2014). Functional information characteristics of large-scale research collaboration: Network measures and implications. *Scientometrics, 102*(2), 1207–1239. doi:10.1007/s11192-014-1475-4.

Heymann, S. (2014). Gephi. In R. Alhajj & J. Rokne (Eds.), *Encyclopedia of social network analysis and mining* (pp. 612–625). New York: Springer.

Himmelstein, D. S., & Powell, K. (2016). Analysis for "the history of publishing delays" blog post v1.0. *Zenodo,*. doi:10.5281/zenodo.45516.

Hunt, J. D., Whipple, E. C., & McGowan, J. J. (2012). Use of social network analysis tools to validate a resources infrastructure for interinstitutional translational research: A case study. *Journal of the Medical Library Association, 100*(1), 48–54. doi:10.3163/1536-5050.100.1.009.

Kolaczyk, E. D., & Csardi, G. (2014). *Statistical analysis of network data with R* (Vol. 65). New York: Springer.

Kumar, S. (2015). Efficacy of a giant component in co-authorship networks: Evidence from a Southeast Asian dataset in economics. *Aslib Journal of Information Management, 68*(1), 19–32. doi:10.1108/AJIM-12-2014-0172.

Larivière, V., Gingras, Y., Sugimoto, C. R., & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology, 66*(7), 1323–1332. doi:10.1002/asi.23266.

Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation, 11*(1), 3–15. doi:10.3152/147154402781776961.

Liu, X., Bollen, J., Nelson, M. L., & Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. *Information Processing and Management, 41*(6), 1462–1480. doi:10.1016/j.ipm.2005.03.012.

Liu, P., & Xia, H. (2015). Structure and evolution of co-authorship network in an interdisciplinary research field. *Scientometrics, 103*(1), 101–134. doi:10.1007/s11192-014-1525-y.

Ministerio de Sanidad y Consumo. Resolución de 30 de marzo de. (2006) del Instituto de Salud Carlos III, por la que se convocan ayudas destinadas a financiar estructuras estables de investigación cooperativa, en el área de biomedicina y ciencias de la salud, en el marco de la iniciativa Ingenio 2010, programa Consolider, acciones CIBER, 83 Boletín Oficial del Estado (pp. 13770–13777).

Newman, M. E. J. (2001a). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E, 64*(1), 016132. doi:10.1103/PhysRevE.64.016132.

Newman, M. E. J. (2001b). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E, 64*(1), 016131. doi:10.1103/PhysRevE.64.016131.

Newman, M. E. J. (2003a). Mixing patterns in networks. *Physical Review E, 67*(2), 026126. doi:10.1103/PhysRevE.67.026126.

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

OECD. (2010). *Measuring innovation: A new perspective*. Paris: OCDE Publishing.

Ramasco, J., & Morris, S. (2006). Social inertia in collaboration networks. *Physical Review E, 73*(1), 016122. doi:10.1103/PhysRevE.73.016122.

Sonnenwald, D. H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology, 41*(1), 643–681. doi:10.1002/aris.2007.1440410121.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036–1039. doi:10.1126/science.1136099.