Co-authorship and academic productivity in economics:

Authors: José Alberto Molina, Alberto Alcolea, Alfredo Ferrer, David Iñiguez, Alejandro Rivero, Gonzalo Ruiz, Alfonso Tarancón

Co-authorship and Academic Productivity in Economics: Interaction Maps from the Complex Networks Approach

José Alberto Molina (1,4,6), Alberto Alcolea (5), Alfredo Ferrer (4,5), David Iñiguez (3,4,5), Alejandro Rivero (4,5) Gonzalo Ruiz(4,5), Alfonso Tarancón (2,4,5)

1. Departamento de Análisis Económico, Facultad de Economía y Empresa, Universidad de Zaragoza, 50005 Zaragoza, Spain

2. Departamento de Física Teórica, Facultad de Ciencias, Universidad de Zaragoza, 50009 Zaragoza, Spain

3. Fundación ARAID, Diputación General de Aragón, Zaragoza, Spain

4. Instituto de Biocomputación y Fisica de Sistemas Complejos (BIFI), Zaragoza, Spain

5. Kampal Data Solutions S.L., 50001 Zaragoza, Spain

6. Institute for the Study of Labor-IZA, Bonn, Germany

Abstract

We explore the relationship between collaborations in writing papers and the academic productivity of economists and, particularly, we describe the magnitude and intensity of co-authorship among economists. To that end, we employ interaction maps from Complex Systems methods to study the global properties of specific networks. We use 8,253 JCR papers from ISI-WOK, published by 5,188 economists from Spanish institutions, and their co-authors, up to 8,202 researchers, from 2002 to 2014, to identify and determine the collaborative structure of economics research in Spain, with its primary communities and figures of influence. Our results indicate that centrality and productivity are correlated, particularly with respect to a local estimator of centrality (page rank), and we provide certain recommendations, such as promoting interactions among highly productive authors who have few co-authors with other researchers in their environment, or recommending that authors who may be well-positioned but minimally productive strive to improve their productivity.

I. INTRODUCTION

The literature on the evaluation of the productivity of a researcher or an institution has grown considerably in the past two decades. Specifically, the excellence of academics or institutions exerts considerable influence on the criteria employed for career advancement or for the assessment of research developed in such institutions. Thus, the evaluation includes output indicators of performance, resulting in the use of objective measures of the individual scholar's quality (e.g. the total number of publications, the total number of citations, the h-index,…).

Economics is no exception to this phenomenon. The economics literature on the relevance of the evaluation of productivity includes a number of important papers. Liebowitz and Palmer (1984) provide a ranking of economics journals based on their relative influence on the writings of academics, with the indicator to create this ranking being the number of citations. Dusansky and Vernon (1998) rank economics departments in the United States, using the criterion of publication in a set of eight leading journals, for the period 1990-94 inclusive. More recently, Oswald (2007) examines data on citations of articles published in the last 25 years, finding that "it is better to write the best article published in a medium-quality journal than all four of the worst four articles published in an elite journal". Kalaitzidakis et al. (2010) provide a ranking of academic journals in economics using data from 2003 to 2009, in such a way that, for each year, the authors count the number of citations of articles published in the previous ten years. Carrasco and Ruíz-Castillo (2014) study the productivity issues of seven cohorts of economists working in the best 81 departments in the world in 2007, concluding, among other things, that "productivity mobility is clearly equalizing in the youngest three cohorts, and it is clearly disequalizing only in the oldest cohort". Using this, and another sample, Albarrán et al. (2014) investigate a number of specific questions of academic productivity, but focus on three elites of economists on the basis of a quality index that weights the publications of each individual in four journal classes. Hamermesch (2015) uses data sets of the post-publication citation histories of articles published in the Top 5 journals, in the 1970s and in the 2000s to analyse specifically the impact of citations on salaries, concluding that the "profession puts too much weight on publications in a few top journals", rather than on the number of citations.

In this context, the general academic literature also shows that collaboration in writing papers confers advantages on co-authors, in such a way that, initially, multi-authored papers generate more citations than single-authored papers, given that both discussion among authors and their varied skills makes a given paper technically stronger (Inzelt et al., 2009; Padial et al. 2010; Researchage, 2013; Schubert, 2014). This is well-documented in Economics, with a clear increase in collaborations among authors over the last years (e.g., Suffer and Kocher, 2004; Goyal et al., 2006; Nowell and Grijalva, 2011: Card and DellaVigna, 2013; Rath and Wohlrabe, 2015), although new results indicates that "citations are not related to the number of authors, which provides evidence against the proposition that multi-authored papers are of higher quality than single-authored papers" (Moosa, 2016).

Despite this evidence, the usual individual measures of academic productivity present certain limitations derived from the complexity of the system that represents the total amount of research in a field or institution. In particular, individual indicators do not capture the interactions between the authors who cooperate with the aim of maximizing the impact and academic prestige that the individual brings to the field or institution. The importance of studying

interrelations appears when we extend the purpose of evaluating individual productivity to a more global concept, analysing cooperation links, and observing what groupings of researchers emerge, how they evolve, and so on.

In this context, the exploration of interaction and collaboration first requires the use of large databases and, secondly, of tools from Complex Networks methods to globally analyse the information. In consequence, the numerical results, primarily graphs and maps, are obtained representing the interactive global information. This information will allow us to identify the targets of possible public or private funds aimed at improving the academic level of researchers, the organisation of research structures and, in this way, facilitating the knowledge of theoretical and empirical, national and international, academic and socio-economic relationships.

The study of collaborative academic networks has recently emerged as an informative application derived from Physics and Computational sciences (see, e.g., Price, 1965, and Cobo et al., 2011, as seminal papers that use Complex Networks and Science Mapping, respectively, to study scientific productivity). In general, when we use data from individuals or aggregated data for Institutions, we are making a local analysis. By contrast, when we make a network study, we are considering global, or topological, properties and, apart from questions such as "who is the author, or group, with the largest number of papers", we emphasise questions like "who is(are) the researcher(s) with more relevance, a greater central role, more cohesion, etc."

In Economics, the international evidence previously cited is based on employing various performance measures from the literature, such as publications in EconLit or in Journal of Citation Reports databases, or as the citations in its multiple versions (total citations, non-autocitations, citations weighted by journal ranking...), but all this evidence presents rankings of individual items (researchers or institutions), with no information about possible interrelations.

To bridge this gap, we develop in this paper network analyses of economists affiliated with Spanish institutions (or publishing with them) in order to represent their interrelations or collaborations, with the final aim of evaluating co-authorship structures, identifying natural communities, hubs, and leaders in the sense of centrality and influence (Newman, 2010), more collaborative regions, deficiencies, … and, finally, detecting whether co-authorship is an indicator of excellence. To perform some of the analyses and to interactively see the results of our study, we use the tools developed by Kampal Data Solutions, located at http://research.kampal.com (Alvarez et al., 2015).

To analyse the productivity and, specifically, the excellence of researchers, one important criterion is how to assign a score to a publication. Here, we use the articles published by researchers affiliated with Spanish academic institutions in the Economics section of the Journal of Citation Reports, and we assign to each paper the JCR Impact Factor (IF) of the corresponding review in the year of the publication. In order to generate informative and comprehensible interaction analyses, a considerable amount of work is needed to obtain and 'clean' all the information, that is, the identification of authors is not automatic in general, and there are bugs due to varying details in signatures and affiliations.

Thus, to construct the network, we define nodes as individual researchers, and links as relations between them generated by common articles. Network analysis will then allow us to see global

properties, as cited above, and the results will be represented by maps that encode the relevant information in graphical terms, as well as numerical parameters that encode the topological properties of the network.

The organization of the rest of this paper is as follows. Section II describes the data scope and processing. Section III describes the construction of networks. Section IV shows the interactive maps and results and, finally, Section V is devoted to our conclusions.

II. DATA

A. Data Scope

We use the ISI-WOK database, selecting publications in the Economics field, from 2002 to 2014, in which at least one author has a Spanish affiliation (i.e. at least one of the institutions he belongs to is in Spain). The total number of papers is 8,253, which have been authored by 8,202 different researchers from 78 countries. We have 5,188 researchers affiliated with Spain.

Within this period, the average annual number of papers published by those authors with Spanish affiliation in this field, is 634.84. This number has grown from around three hundred in 2002 to about one thousand from 2011 on. The publications have been extracted from 327 distinct journals, the average impact factor of the articles is 0.98, each paper is cited 7.2 times, on average, and the eight countries that collaborate most often with Spain in the field are (in decreasing order) the USA, England, Germany, Italy, the Netherlands, France, Belgium, and Canada.

B. Names and Affiliation Filtering

For each selected paper, we obtain a list of all the author names, with the corresponding affiliations. In general, names and affiliations present small (though occasionally large) differences for the same author in different papers.

To obtain better results, we perform a process of unification. The most important questions we address in this process have to do with the following problems. First, the same author can use a significant number of different signatures (first name, last name, a rearrangement of them, special characters…) and, second, the same author can use different ways to specify his affiliation (address, centre, city…). Additionally, one author can, in some cases, change his/her affiliation, corresponding to an actual relocation from one institution to another.

To minimise the impact of these circumstances, and to obtain realistic results, we reduce the original number of different researchers by unifying some of their identities. To ensure the results are correct, we run a large number of tests and crosschecks, and we manually compare names and institutions to detect differences, and to associate the same researcher with different signatures and/or affiliations. To detect that two different signatures or addresses are the same, we use the Levenshtein distance between strings (Levenshtein, 1966), where distance is defined as the number of insertions or deletions needed to convert one string into another.

Taking these factors into account, we consider the following criteria in order to identify authors as accurately as possible:

1. Order of first name and last name

2. Levenshtein distance between first name and last name, smaller than 5%

3. Initials

4. Same city of affiliation

After running this analysis, we found that the papers were signed by a total of 8,202 different authors (including those from countries other than Spain) and the total number of authors from Spanish institutions was 5,188.

C. Merits Not Present in the Study

To check the reliability of our tracking, we select a small group of authors and run a detailed search of their scientific production, finding significant differences in some cases, due to one or more of the following:

- Articles by researchers who are currently affiliated in Spain, but that at the moment of the publication were affiliated in another country

- Articles written before 2002

- Articles published in journals that do not belong to the field of economics (but to areas such as physics or mathematics)

- Names with very different typography

- Publication in journals not indexed in JCR

Thus, occasionally, the number of articles in our analysis is smaller than in the full CV of the researcher, but the selected subset corresponds adequately to the specifications of our study.

III. NETWORKS

We construct and analyse the network of economists affiliated with Spanish academic institutions by applying tools and methodologies from the Complex Networks discipline (Boccaletti et al., 2006). In this way, we are able to represent these associations as maps or graphs (graphical representations of the networks) that allow us to focus on the structural and cooperative aspects of the global structure. A detailed description of the procedures to construct academic networks can be found in Alvarez et al. (2015).

To build the Network, we must define nodes and links, and we begin by selecting the nodes (researchers) we want to study. There are many possible criteria to do this, taking into account, for instance, the field of work (through the selection of a certain subset of reviews, as we have done here) of the affiliation of the author, which allows us to limit the analysis to a certain institution or some other geographical aggregation (such as city, region, or country). Other criteria could be used, based for example on the excellence of the publications, as we will do in one of our case studies at the end of this paper.

Next, we define the links that join the nodes. We link two authors if they have at least one JCR article signed in common in the period under study. We then assign a weight to each link according to the Impact Factor (IF). Thus, the weight associated with the link between two authors produced by the common publication of an article in a certain journal is the JCR Impact Factor of that journal, in the year the article was published, divided by its total number of co-authors (number of authors minus one). The weights depend on the journal of publication of the paper, not on the details of the paper. All papers from the same review in the same year have exactly the same weight. If two authors share more than one publication, we sum the weights of all the corresponding links in order to obtain the total weight or strength of their relationship. Additionally, we also assign a weight to each link according to an Excellence (EXC) criterion, within which we only consider papers in the first decile, all of those with weight 1 (also divided by the number of co-authors).

Here, the decision to divide by the number of co-authors is not to give more importance to the papers with a smaller number of authors. On the contrary, we do this because when we sum over all the links generated by a particular paper for a certain author (node), we recover the corresponding value of IF or EXC (because the number of links coincides with the number of co-authors, i.e. number of authors minus 1). In fact, in order to prioritise the papers with a smaller number of authors, one should divide again the previously defined weights, in this case by the number of authors of the paper (including the node itself). We shall call such factors "author normalised (AN) weights" or, shortly, AN-IF and AN-EXC weights, respectively.

Once we have defined the scope of the study through the selection of the nodes (authors), and the strength of the relationship between each pair of authors through this procedure, we obtain the corresponding networks of collaborations between researchers. Once a particular network has been constructed, we can study a large number of properties, such as cluster formation, topology, diameter, mean distance, assortativity, etc. These measurements quantify the topological properties of the network (see Boccaletti, 2006, for Complex Systems and, particularly, Alvarez et al., 2015, for details of Research Based Networks).

A. Degree

For each particular node, we have many links joining it to all collaborators, each one having a weight calculated as explained above. Then, the degree of a node is computed by adding the weight of all the links that join a node. In order to calculate the weight of each individual link, we divide the original impact by the number of co-authors of the corresponding article as explained above. Thus, the degree is the total impact of the papers published by this node (author), except for the following detail.

To be more precise, we add the links present in the map and, therefore, the total corresponds to the activity between researchers present in the same map. In other words, if one author collaborates only with people in the map, his/her degree corresponds to all his/her merits; by contrast, if a researcher has external links to collaborators not present in the map, this part of the merit is lost in the computation of the degree (i.e. if an author has an article with a non-Spanish-affiliated researcher, his/her merit in the network is 50% of the corresponding IF weight, while if the collaboration is with a Spanish-affiliated researcher, the merit is 100% of the weight). We do this to show the actual internal activity of the network, and the relative importance of each node within its community, independent of the total scientific production in other groups, which, of course, is fully valid as an individual measure of academic productivity.

When working with the original weights (IF, EXC), we obtain, in fact, that the degree is the total impact of the papers published by this author (inside the network under study, due to the reason we have just explained). For instance, consider a paper with IF=2 in the first Quartile published by 4 authors. The network associated with this paper consists of 4 nodes, each with 3 links connecting to the other nodes. If we consider the IF weight, each link has a value of 2/3. Now we can compute the degree associated with a node in each case, by summing over the 3 associated links. In the IF case, the result is 2, with this number effectively corresponding to the total merits of the author, that is to say, the node, generated by the specific paper.

However, if we sum the degree of the four authors of the article, this paper will contribute 4 times its impact to the total degree of the network, while an article with a single author will contribute only once. This anomaly is solved when using the AN-weights explained above, though both kinds of weight can be valid.

B. Geometry and Graphical Representation

One of the more interesting and self-explanatory results is the graphical representation of the network (a positions map). For that, we run positioning algorithms to allocate nodes closer to their more collaborative colleagues. We use force-directed algorithms (Fruchterman and Reingold, 1991) and, additionally, a Monte Carlo process to separate overlapping researchers.

The procedure is as follows. We define a set of interacting particles (one for each node of the network), which evolve under the effect of Newton's laws. Particles are represented by coloured discs of area proportional to the degree of the node they represent. The force encoding the interaction between two particles corresponds to the weight of the link connecting both nodes in the corresponding network (as does the thickness of the line representing it).

To improve the presentation, we include certain technical details, such as a short-range force that prevents the nodes getting too close to each other, assuming the interaction to depend only on the distance between the nodes in the plane, and fixing a global scale for the map to scale all the distances and magnitudes of the discs.

We let the system evolve under these interactions, and obtain (by means of standard algorithms of statistical mechanics) a stable equilibrium position for the system.

C. Communities

Based on the attraction forces we have just explained, those algorithms assign a position to each node in such a way that nodes with more collaborations between them are, in general, closer, which is not easy because the structure of relationships is complex. Finally, the system arrives at a situation with relatively compact groups (communities), where internal collaboration is strong, linked to the rest through some connections weaker than the internal ones. We can draw an analogy to a galaxy, where every star is attracted to each other, but in the end they form clusters in the regions of higher density, and every cluster differs from the rest.

In order to detect this communities structure, we apply a few standard algorithms of clustering, to identify those nodes with a stronger interaction among them. We use different algorithms, because their performance depends on the size and on the directed-undirected character of the map. We mainly use walktrap (Pons and Latapy, 2006) or leading-eigenvector algorithms (Newman, 2006). Each subgraph constitutes a community in the network, and their nodes and links can be coloured in the same way (and differently from the other communities). Thus, we see at first glance the partnership structure of the researchers under study.

D. Centralities

A natural question that emerges from a collaborative network is which are the most cohesive nodes, the most important, or those with the greatest authority. We can define different kinds of centrality measures to quantify these aspects (Newman, 2006). For example, betweenness (BTW) indicates the importance of a node in communicating with different communities, and the nodes with the largest values are the most appropriate, for instance, to host activities tending to interdisciplinary collaboration, while Page Rank (PRK) centrality deals with the number of important nodes that point to it (Ying et al., 2010).

Centrality measures give us different information from production. Production (for example, number of papers, or IF), is a local quantity, which does not consider collaborative and global aspects. On the contrary, centrality tells us how the researcher produces papers, what is their position in the group, their importance in maintaining collaborations, and their relevance or authority.

D. Global Network Properties

Now, let us introduce briefly a set of variables describing the important properties of a network, which we will refer to below.

- Giant Cluster Size: Number of nodes of the largest cluster of the network.

- Average Path Length: Average number of (weighted) steps along the shortest paths for all possible pairs of network nodes. This number conveys (on average) how many co-authors we must traverse to move from one to any other researcher on the map.

- Assortativity: Preference for the nodes to attach to others that are similar in some way; for example, the tendency of highly-connected nodes to be connected to other high-degree nodes. Positive values show that the main researchers collaborate among

themselves, while negative values indicate that the main researchers only collaborate with colleagues without important roles.

- Clustering: Probability that two neighbours of a third party are at the same time neighbours of each other.

- Modularity: Presence of well-defined subgroups or clusters. In general, values above 0.4 correspond to networks with a large number of small clusters with high internal communication.

IV. RESULTS

All the network properties defined previously, such as degree, geometry, centralities, etc. depend on the weight used and on the set of researchers selected. In this paper, we only show some of the possibilities, while the rest can be seen in the online complementary information on http://research.kampal.com/visualization/economics-research-spain/.

As explained above, the scope of the study is defined by selecting the ISI-WOK database of publications in the Economics field from 2002 to 2014, in which at least one author has a Spanish affiliation (i.e. at least one of the institutions he belongs to is in Spain). The total number of papers is 8,253, authored by 8,202 different researchers from 78 countries. Of course, most of them (5,188) belong to Spanish institutions, but we have also considered the rest of the authors (only through the papers included in this selection).

A. Global Maps and General Results

Working with the global map containing all the researchers in that scope, we construct the map of Figure 1 where, as explained above, two nodes are connected by links if both researchers have authored a certain paper. The size of each researcher is computed as the sum of the AN-IF weight for all his selected papers.

 (Figure 1 about here)

This map has thousands of nodes, and thus it is difficult to appreciate the details (the online complementary information allows a more detailed visualisation of the map, together with all the analysed data, maps, and results). Even so, one can see a large region around the centre of the figure where all the researchers are, directly or indirectly, connected (i.e. the giant cluster), while on the periphery one can find nodes with null or very scarce collaborations.

On this map we can automatically identify communities (groups of researchers with stronger collaborations), and show them in such a way that all researchers from the same detected community have the same colour. A detailed inspection allows us to see several regions corresponding to different authors; e.g., the dark area of the upper central zone at the top of the Figure corresponds to van den Bergh, JCJM (UA, Barcelona), whereas the green area on the central-east zone corresponds to Galí, J (UPF, Barcelona).

The values of some of the network properties are as follows.

- Average AN-IF of the researchers: 0.95 (while the maximum value is 27.75, which indicates a rather large dispersion)

- Giant Cluster Size: 4,830 (of the total 8,202 nodes of the network)

- Average Path Length: 8.96 (which is a rather large value due to the fact that the network is not very dense)

- Assortativity: 0.77 (which indicates that the collaboration is not very hierarchical)

- Clustering: 0.57 (a rather cohesive network)

- Modularity: 0.94 (it is a very modular network, with intense collaborations within the modules and scarce intermodular connections)

It is interesting to compare the previous map with an Excellence Map, i.e. the map constructed only with papers of the first JCR decile, and using the AN-EXC weight (Figure 2). We can see, as expected, a less-populated and less-connected map than in the previous case. For example, the beige area in the central-east of the Figure corresponds to van den Bergh, JCJM (UA, Barcelona) and the dark green area in the upper-west corresponds to Voth, HJ (UPF, Barcelona).

(Figure 2 about here)

We have asked ourselves whether the Spanish authors produce papers mainly between themselves, or does there exist an important collaboration with foreign researchers, and if this grade of collaboration depends on the importance of the author.

To compare total (number of papers of an author) versus internal (fraction of papers produced by Spanish authors) production, we represent in Figure 3 both numbers on a scatter plot, and we compute the Pearson coefficient of correlation, which turns out to be 0.9813, showing that international collaborations are almost proportional to the scientific production without, in general, very significant differences between researchers.

(Figure 3 about here)

In the next sections, we inspect the results in detail, showing different possibilities of describing the importance of the researchers, as well as comparisons between different modes of cooperation.

B. Production and Centralities

In order to give an estimation of the importance of the different authors, one could use different magnitudes. Of course, one of these is the degree that, in the way it is defined, it is directly related to the volume and impact of the publications. However, one can also use other kind of magnitudes, more associated with a network approach, such as some of the possible definitions of centrality. Here, we have chosen the betweenness and page rank described above.

The production and centrality estimators can be rather different for the same author, as the first measures individual production, while the other indicates the importance in terms of the relations with the rest of the researchers (though of course they will be partially correlated).

In Table 1, we show the value of these variables for the top 10 authors in terms of production. We have added a column with the number of co-authors because it is useful in order to understand the structure of the collaborations established by each investigator.

While in terms of production it is difficult to find extremely large differences (because, with rare exceptions, the capacity to generate papers for a researcher is finite, and the number of publications of an author grows rather linearly with the years of work), the betweenness factor can show differences of several orders of magnitude between central and peripheral researchers.

 (Table 1 about here)

We can see that van den Bergh, JCJM (UA, Barcelona), Galí, J (UPF, Barcelona), and Voth, HJ (UPF, Barcelona) appear in the former positions. Additionally, some of the most productive researchers also have higher values of centrality (BTW and/or PRK), with some particular cases of lower centrality, as is the case of Voth, HJ (UPF, Barcelona) because all his publications are with non-Spanish authors.

With regard to the two estimators of centrality, we can see that some authors have large values of both, e.g., Loureiro, ML (USC, Santiago de Compostela) and Gil-Alaña, LA (UN, Pamplona), among others, indicating that they have a rather central position both in relation to the whole network (betweenness) and to their region of influence (page rank), while others, such as Galí, J (UPF, Barcelona) have a smaller value of BTW, despite that they are very powerful leaders in their fields (large PRK value).

It is also interesting to show the Excellence Table, i.e. the table constructed only with papers on the first JCR decile and using the AN-EXC weight (Table 2). We can see similarities with respect to the upper-half of the Table, with the same authors appearing in some of the former positions, that is to say, van den Bergh, JCJM (UA, Barcelona), Voth, HJ (UPF, Barcelona), Galí, J (UPF, Barcelona) and Vives, X. (IESE, Pamplona).

 (Table 2)

On the other hand, if we plot the histogram of the number of researchers who have a certain IF value (Figure 4), these top 10 researchers correspond to the right-side region of the figure, which could be denoted as the Excellence region. In this figure, we can see that the great majority of researchers have a very small IF value (note the logarithmic scale on the Y axis) and that the number of authors decreases very rapidly when growing the IF on the X axis (see Clauset, 2009 for more information about the expected decay), until reaching the excellence region.

(Figure 4 about here)

We can repeat the previous analysis by considering only papers on the first decile (Excellence). The histogram of the number of researchers who have a certain number of papers in the first IF decile has the aspect shown in Figure 5. The appearance is similar to the total IF histogram, but

now the large concentration of people in the smaller values of the production is much more evident, with the great majority of researchers having no articles in that first decile.

(Figure 5 about here)

As a possible indicator of the influence of the network of collaborators on the productivity of a researcher, we have asked ourselves whether centrality and production are correlated. At first sight, in the tables above, it seems that there exists some correlation, which is more evident when we take a more local estimator of centrality (page rank) than when we use a more global measure (betweenness). In fact, the Pearson coefficient between IF and PRK is 0.8321, while between IF and BTW it is 0.5886. The scatter plot in Figure 6 illustrates this relationship.

(Figure 6 about here)

However, this plot is not very precise if we want to study whether there exist, in the IF-BTW *phase space*, regions representing the different behaviours of researchers. To go a little deeper into this, we compute the mean value and dispersion (standard deviation) of each of these two variables (we use the logarithm in order to highlight the differences between researchers in the low-productivity region) and then we calculate the distance from the mean value of every researcher, in standard deviation units (for each variable). Figure 7 is the histogram of the number of authors in each bidimensional interval. In this Figure, the X axis corresponds to the distance for the IF variable, the Y axis is the distance for BTW, and the Z axis (number of researchers) is substituted by a colour code, going from clear (larger values) to dark (smaller values). As the distributions are not Gaussian (in fact they are very different), the mean is different from the median and the mode. We can see several regions corresponding to authors who are well-positioned both in IF and BTW (upper right region) with representatives such as Gil-Alaña, LA. (UN, Pamplona), others with a large IF but small BTW, such as Voth, HJ (UPF, Barcelona), or others with large BTW but a bit lower IF, such as Loureiro, ML (USC, Santiago de Compostela). On the upper left region of the Figure, we can find authors who, despite being well-positioned in the network, occupying a rather central position or connect different communities, do not have a good ratio of publications.

This indicates that, despite that the correlation between publications and centrality is quite important, the relationship is not trivial. In fact, from the perspective of an R&D manager, knowing these values for a particular researcher could provide certain recommendations, such as promoting interactions between highly productive authors who have few collaborations with other researchers in their environment, or asking authors who may be well-positioned but minimally productive, to strive to improve their production.

(Figure 7 about here)

C. Collaborative Structure and Communities

In this subsection, we study in detail the structure of the collaborations established between researchers. We begin by analysing the typical number of collaborators of an author. Figure 8 shows the histogram of the number of authors who have a certain number of co-authors (for each X value, we plot the number of researchers who have X collaborators in the Spanish network). We can see that the most usual number of co-authors is zero, with nearly 25% of the cases (note the logarithmic scale on the Y axis) and then the number of occurrences decreases monotonically with the number of collaborators following a power law, approximately. The number of authors who have a significant number of collaborators is not large, implying that the network is not very dense, as we mentioned earlier, but it is sufficient to have a rather consistent network and to find a giant cluster to which more than half of the researchers belong.

(Figure 8 about here)

From these collaborations, established among the different researchers, we are interested in what kind of groups or collaborative structures emerge. We try to determine this by studying particular cases, through two different types of analysis. First, we use automatic algorithms to detect the communities that appear naturally from the established collaborations, as well as the authors belonging to each. Second, we prune the network around certain researchers, limiting it to first- or second-neighbour distance. In the first case are the algorithms that, without our intervention, decide which researchers conform to a certain community of collaboration subnetwork, while, in the second case, we *define* an author's working environment as the direct (or almost direct) collaborators.

Both kinds of analysis are possible, thanks to the use of complex network techniques, with this being one of the differential values we can obtain in comparison to local (individual) statistical analysis. For the automatic detection, we have used the algorithms cited above and we have obtained a series of communities, from which we show, in Table 3, the larger ones according to their total production. For each, we present in the table the most relevant author in terms of AN-IF, which we have called the community leader, the total AN-IF of the community, the AN-IF internal to the own community (i.e. not including the part of the AN-IF provided by the collaborators outside the community) and the number of nodes (researchers) that form it. Here, we observe the presence in the former position of Martin, JC. (ULP, Las Palmas).

(Table 3 about here)

We can do a similar analysis constraining the network to be formed only by the authors affiliated with Spanish institutions, which we have called the "Spanish ecosystem" (i.e. not including their foreign collaborators). In this case, we obtain the communities shown in Table 4. We can see similarities with respect to the upper-half of the Table, with the same authors appearing in the former positions, that is to say, Calvo-Armengol, A. (BGSE, Barcelona) and Martín, JC. (ULP, Las Palmas).

 (Table 4 about here)

In order to illustrate the structure and components of some of these communities, we have selected those of Calvo-Armengol, A. (BGSE, Barcelona) and Galí, J. (UPF, Barcelona)

corresponding to the first and last positions in our Top 10 Spanish Communities Table. We present the 10 most relevant members of each in terms of AN-IF in Tables 5 and 6. We can see that, in the first case there are no great differences between the production of one and other members, while in the second, the leadership of Galí, J. (UPF, Barcelona) is clearer.

(Table 5 about here)

(Table 6 about here)

Another way to deepen this line of analysis is the study of the networks formed up to first or second neighbours distance of a researcher, as cited above. This is a different approach to the previous one, where the communities were automatically detected by the analytical complex networks tools. If we take, for example, the second neighbours network formed around the economist Calvo-Armengol, A. (BGSE, Barcelona), we obtain the map shown in Figure 9, while the equivalent map for Galí, J. (UPF, Barcelona) has the aspect presented in Figure 10. Again, in the first case, we can see a rather uniform network, while the second presents a more hierarchical structure, with Galí, J. (UPF, Barcelona) having an important pre-eminence over the rest of the researchers, especially over his direct collaborators.

(Figure 9 about here)

(Figure 10 about here)

V. CONCLUSIONS

We have analysed the production and structure of the collaborations among a set of researchers, economists from Spanish institutions, from a complex networks perspective. One can define network-related variables providing information that is complementary to the classical individual rankings of academic productivity. We have observed that centrality (a measure of the connectivity, or good positioning, of each author in the collaborative network) has a positive correlation with production and that, studying both variables simultaneously, one can obtain interesting conclusions that could be used for the improvement of research from a global perspective. Our results and the complementary information appear on http://research.kampal.com/visualization/economics-research-spain/, in this way allowing for a more detailed visualisation of the maps, together with all the analysed data and results.

Results on production using total JCR impact indicate van den Bergh, JCJM (UA, Barcelona), Galí, J (UPF, Barcelona) and Voth, HJ (UPF, Barcelona) appear in the former positions, with these positions also appearing for the excellence JCR impact. Regarding the two estimators of centrality, we can see that certain authors have large values of both, e.g., Loureiro, ML (USC, Santiago de Compostela) and Gil-Alaña, LA (UN, Pamplona), among others, indicating that they have a rather central position both in relation to the whole network (betweenness) and to their region of influence (page rank).

The analysis of the topology of the collaborations shows the existence of communities with large internal connectivity and with a rich internal structure. With respect to the relevance of these communities, we highlight the presence in the former position of Martin, JC. (ULP, Las Palmas), and if we constrain the network to be formed only by the authors affiliated with Spanish institutions, we can see similarities, with the same authors appearing in the former positions, that is to say, Calvo-Armengol, A. (BGSE, Barcelona) and Martín, JC. (ULP, Las Palmas).

In sum, our analysis indicates that whether centrality and production are correlated, it is more evident when we take a more local estimator of centrality (page rank) than when we use a more global measure (betweenness). This could provide certain recommendations for the future, such as promoting interactions between highly-productive authors who have few collaborations with other researchers in their environment, or encouraging those authors who may be well-positioned but minimally productive to strive to improve their production.

REFERENCES

Alvarez, R., E. Cahué, J. Clemente-gallardo, A. Ferrer, D. Iñíguez, X. Mellado, A. Rivero, G. Ruiz, F. Sanz, E. Serrano, A. Tarancón and Y. Vergara. "Analysis of academic productivity based on Complex Networks." *Scientometrics*, 104, 2015, 651 - 672.

Albarrán, P., R. Carrasco and J. Ruíz-Castillo. "The elite in Economics." Mimeo. 2014.

Bergantiños, G., J.M. Da Rocha and P. Polomé. "La investigación española en Economía, 1995-1999." *Investigaciones Económicas*, 26, 2002, 373-392.

Boccaletti, S., V. Latora, Y. Moreno, M. Chavez and D. U. Hwang. "Complex Networks: Structure and Dynamics." *Physics Reports*, 424, 2006, 175-308.

Card, D. and S. DellaVigna. "Nine facts about top journals in Economics." Journal of Economic Literature, 51 (19), 2013, 144-161.

Carrasco, R., and J. Ruiz-Castillo. "The evolution of the scientific productivity of highly productive economists." *Economic Inquiry*, 52, 2014, 1-16.

Clauset A., C.R. Shalizi and M.E.J. Newman. "Power law distributions in empirical data". *SIAM Review*, 51, 2009, 661-703.

Cobo, M.J., A.G. López-Herrera, E. Herrera-Viedma and F. Herrera. "Science mapping software tools: review, analysis, and cooperative study among tools." *Journal of the American Society for Information Science and Technology*, 62 (7), 2011, 1382-1402.

Dusansky, R., and C.J. Vernon. "Rankings of U.S. Economics Departments." *Journal of Economic Perspectives*, 12, 1998, 157-170.

Fruchterman, T.M.J. and E.M. Reingold. "Graph Drawing by Force-directed." *Software: Practice and Experience*, 21(11), 1991, 1129.

Goyal, S., M.J. Van Der Leij and J.L. Moraga-González. "Economics: an emerging small world." *Journal of Political Economy*, 114 (2), 2006, 403-412.

Hamermesh, D.S. "Citations in Economics: measurement, uses and impacts." IZA DP No. 9593, 2015.

Inzelt, A., A. Schubert and M. Schubert. "Incremental citation impact due to international co-authorship in Hugarian higher education institutions." *Scientometrics*, 78, 2009, 37-43.

Kalaitzideakis, P., T. P. Mamuneas and T. Stengos. "An updated ranking of academic journals in Economics." The Rimini Centre for Economic Analysis WP 10-15, 2010.

Liebowitz, S.J., and J. P. Palmer. "Assessing the relative impacts of economic journals." *Journal of Economic Literature*, 22, 1984, 77-88.

Levenshtein, I. "Binary codes capable of correcting deletions, insertions and reversals." *Cybernetics and Control Theory*, 10, 1966, 7076710.

Moosa, I.A. "Citations, journal ranking and multiple authorships: evidence based on the top 300 papers in Economics." *Applied Economics Letters*, DOI:10.1080/13504851.2016.1176104.

Newman, M.E.J. "Finding community structure in networks using the eigenvectors of matrices", *Physical Review E*, 74(3), 2006, 036104.

Newman M.E.J. *Networks: An Introduction*. Oxford University Press, 2010.

Nowell, C. and T. Grijalva. "Trends in co-authorship in Economics since 1985." *Applied Economics*, 43 (28), 2011, 4359-4375.

Oswald, A.J. "An examination of the reliability of prestigious scholary journals: evidence and implications for decision-makers." *Economica*, 74, 2007, 21-31.

Padial, A.A., J.C. Nabout, t. Wiqueira, L.M. Bini and J.A.F. Diniz-Filho. "Weak evidence for determinants of citation frequency in ecological articles." *Scientometrics*, 85, 2010, 1-12.

Pons, P. and M. Latapy. "Computing communities in large networks using random walks." *Journal of Graph Algorithms and Applications*, 10 (2), 2006, 191-218.

Price, D.J. de S. "Networks of scientific papers." *Science*, 149 (3683), 1965, 510-515.

Rath, K. and K. Wohlrabe. "Recent trends in co-authorship in Economics: evidence from RePEc." *Applied Economics Letters*, DOI:10.1080/13504851.2015.1119783.

Researchage. "Why is it becoming increasingly rare to see single-authored journal papers." October 31, 2013.

Schubert, A. "Sentences to remember from the first 100 volumes of the Journal Scientometrics." *Scientometrics*, 100, 2014, 1-13.

Sutter, M. and M. Kocher. "Patterns of co-authorship among Economics Departments in the USA." *Applied Economics*, 36 (4), 2004, 327-333.

Ying D., Y. Erjia, A. Frazho and J. Caverlee. "PageRank for ranking authors in co-citation networks." *Journal of the American Society for Information Science and Technology*, 60 (11), 2009, 2229-2243.

FIGURE 1

Global Map

Global Map containing all the authors, using the AN-IF weight. The colour of each node
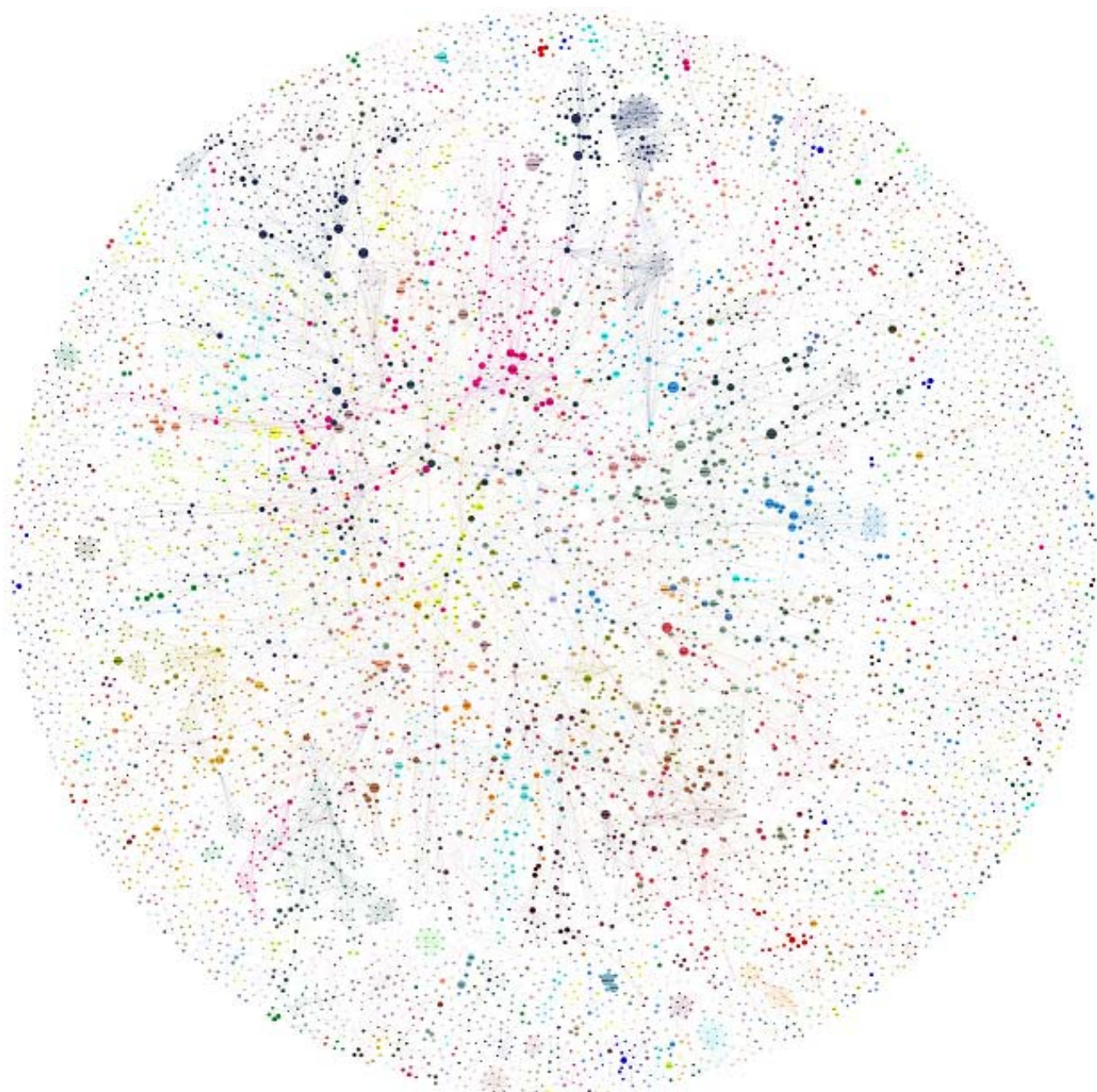corresponds to the automatically-detected community.

FIGURE 2

Excellence Map

Excellence Map containing the authors with at least one paper in the first JCR decile, using the AN-EXC weight. The colour of each node corresponds to the automatically-detected community.
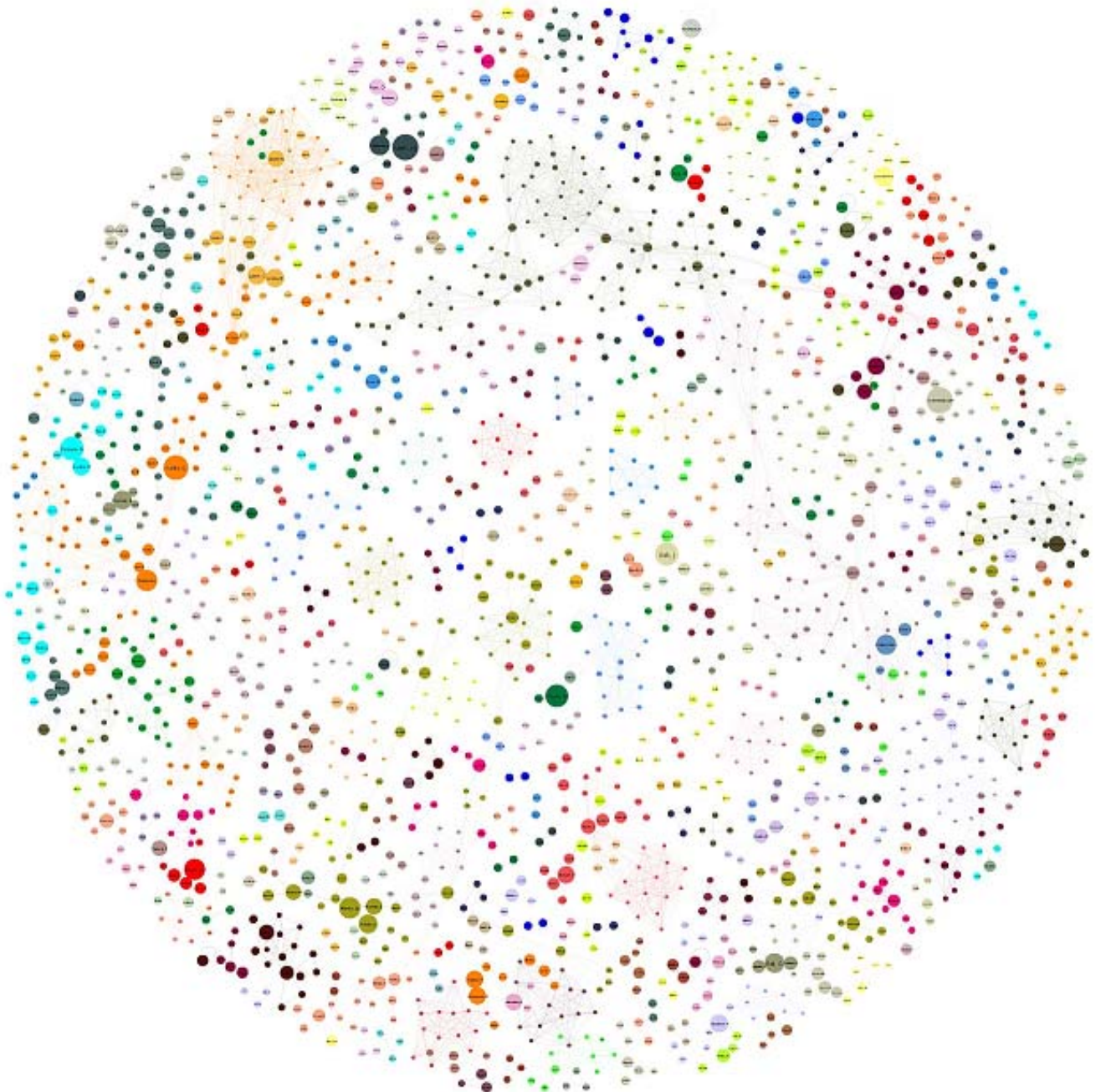
FIGURE 3
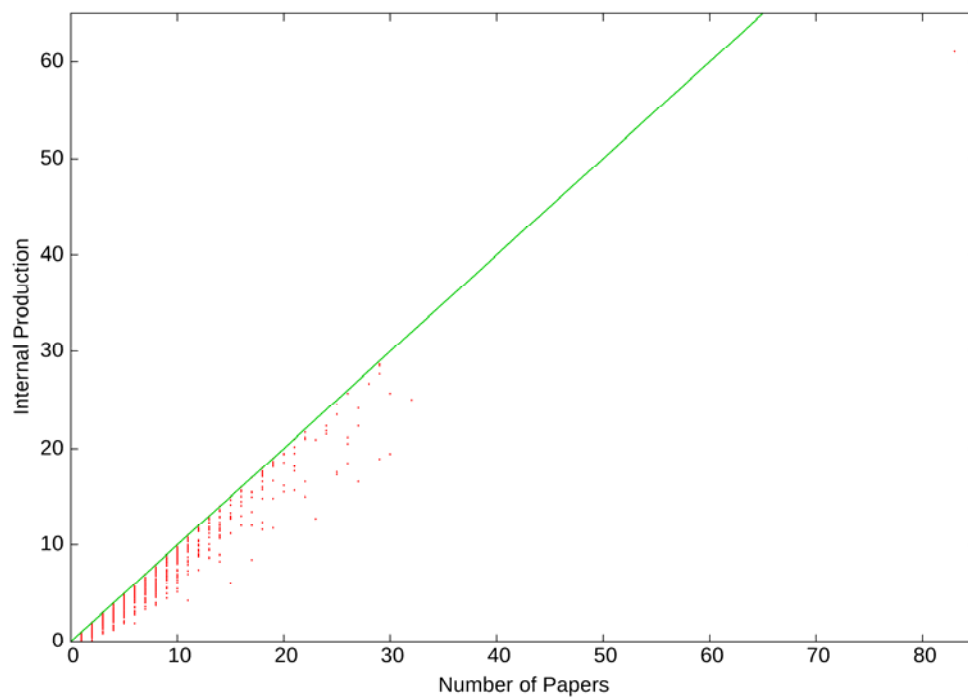
Scatter Plot of Total and Internal Production

TABLE 1

Ranking of the Top 10 Researchers according to their production AN-IF

Name (University)	Production (AN-IF)	Betweenness	Page rank	Co-authors
van den Bergh, JCJM (UA, Barcelona)	27.75	0.1791	0.79076	16
Galí, J (UPF, Barcelona)	27	0.05797	0.95944	20
Voth, HJ (UPF, Barcelona)	26.46	0.00109	0.00109	7
Gil-Alañaa, LA (UN, Pamplona)	21.93	0.26829	1	27
Vives, X (IESE, Pamplona)	20.43	0.01771	0.3966	7
Canova, F (UPC, Barcelona)	17.73	0.21371	0.53063	12
Rodriguez-Pose, A (LSE, London)	16.97	0.20433	0.50273	6
Ciccone, A (BGSE, Barcelona)	16.95	0.24803	0.49595	11
Romero-Avila, D (UPO, Sevilla)	16.4	0.0252	0.47758	6
Loureiro, ML (USC, Santiago de Compostela)	16.09	0.62781	0.73083	23

TABLE 2

Ranking of the Top 10 Researchers according to their production AN-EXC

Name (University)	Production (AN-EXC)	Betweenness	Page rank	Co-authors
van den Bergh, JCJM (UA, Barcelona)	6.08	0.02298	0.02298	8
Voth, HJ (UPF, Barcelona)	6	0.00985	0.00985	4
Kallis, G. (UAB, Barcelona)	5.37	0.66964	0.53215	15
Ciccone, A (BGSE, Barcelona)	4.83	0.04616	0.04859	5
Galí, J. (UPF, Barcelona)	4.67	0.02462	0.0348	3
Vives, X. (IESE, Pamplona)	4.5	0	0.00328	1
Martínez-Alier, J. (UAB, Barcelona)	3.94	0.60388	0.53954	19
Martín, JC. (ULP, Las Palmas)	3.92	0.08376	0.13119	6
Duro, J.A. (UAB, Barcelona)	3.83	0.00948	0.01422	3
Rodriguez-Pose, A (LSE, London)	3.5	0.00821	0.01094	2

FIGURE 4

IF Histogram

For each IF value interval, we plot the number of researchers who have their total impact within it (logarithmic scale on the Y axis)
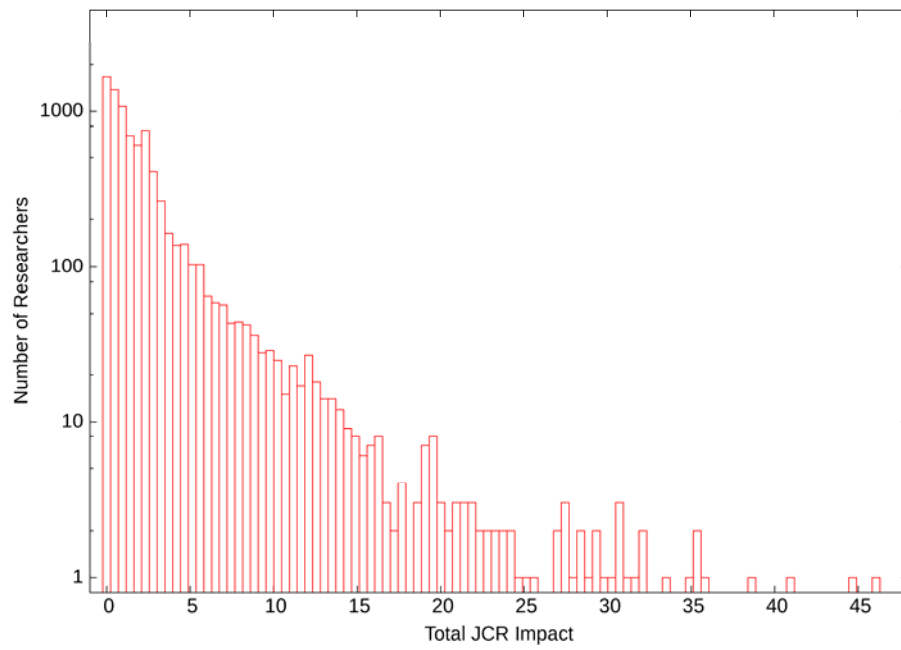
FIGURE 5

Histogram of Excellence Papers

For each value of the number of papers in the first IF decile, we plot the number of researchers who have exactly that number (logarithmic scale on the Y axis).
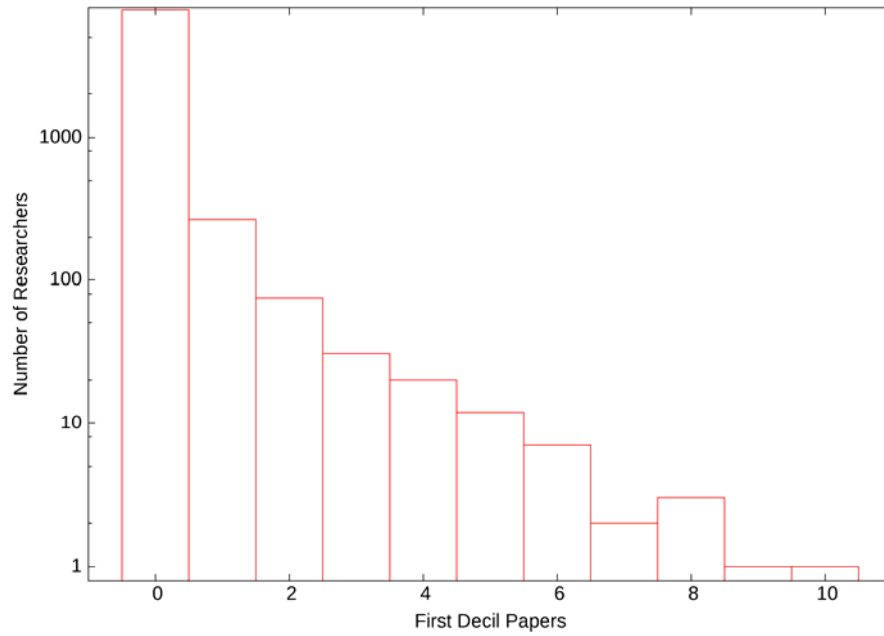
FIGURE 6

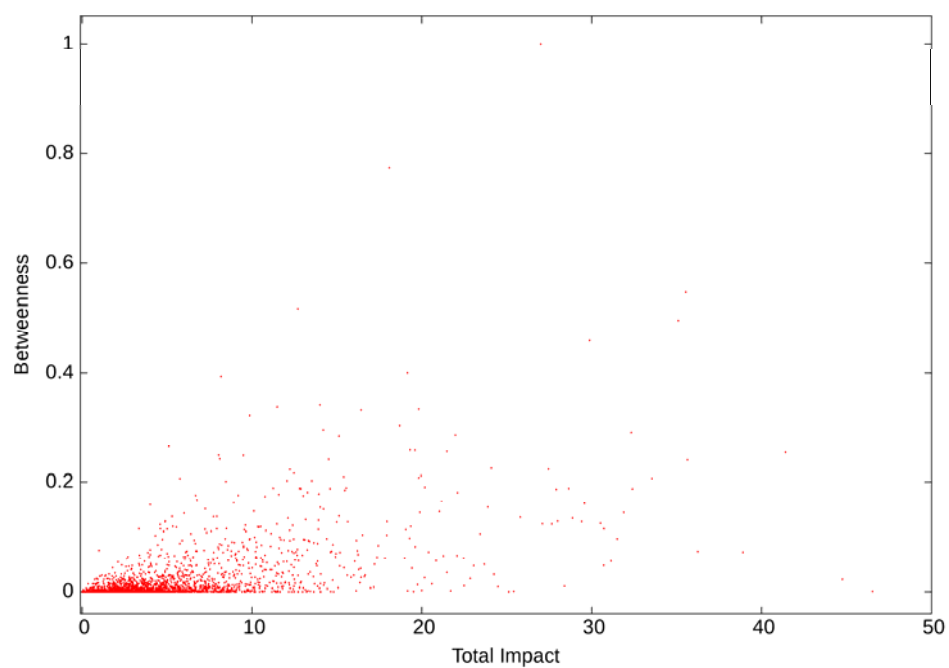Scatter Plot of Impact Factor and Betweenness

FIGURE 7

Colour Map of Number of Researchers as a Function of Impact Factor (X axis) and Betweenness (Y axis).
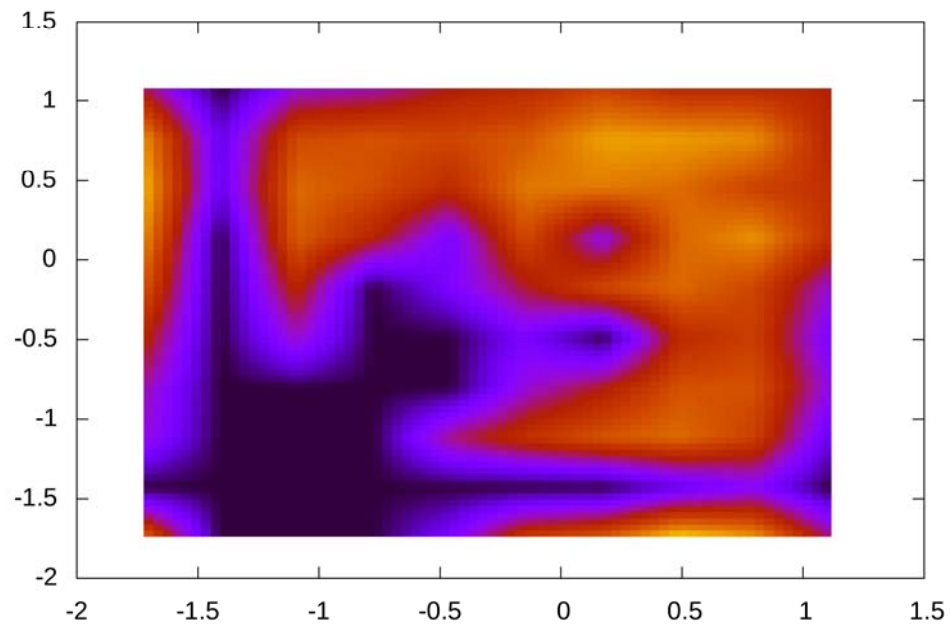
FIGURE 8

Histogram of Number of Co-authors

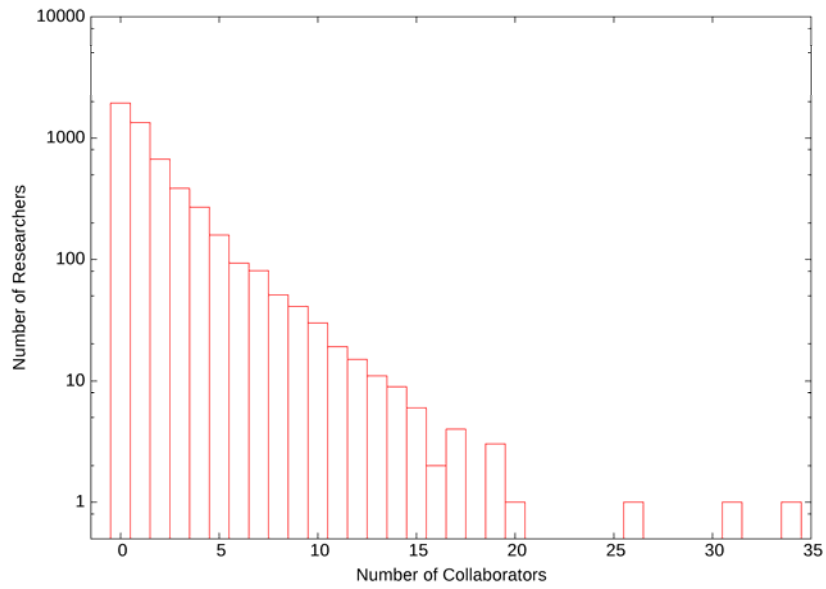For each X value, we plot the number of researchers who have X co-authors

TABLE 3

Top 10 Automatically-Detected Communities (AN-IF weight)

Community Leader	Production in the Network	Production in the community	Number of Researchers
Martin, JC	287.06	278.66	190
Calvo-Armengol, A	204.22	194.79	113
Brandts, J	182.38	177.90	170
Vives, X	166.67	164.64	94
Rodriguez-Pose, A	166.70	164.13	133
Bel, G	163.46	159.78	110
Holl, A	165.61	158.99	155
Tortosa-Ausina, E	163.66	158.78	145
Guner, N	159.68	156.81	122
Galí, J	150.81	146.48	104

TABLE 4

Top 10 Automatically-Detected Communities in the Spanish Network (AN-IF weight)

Community Leader	Production in the Spanish Network	Production in the community	Number of Researchers
Calvo-Armengol, A	192.08	188.80	90
Arrunada, B	168.34	164.33	132
Ezcurra, R	161.00	159.42	93
Martin, JC	140.67	136.70	76
Holl, A	139.47	134.32	109
Leon, C	135.13	132.10	95
Casares, M	132.53	128.39	84
Mora, T	127.11	124.55	81
Brandts, J	108.94	105.93	89
Galí, J	105.85	103.22	68

TABLE 5

Top 10 Members of Calvo-Armengol Community

Researcher	Production
Calvo-Armengol, A	12,89
Montalvo, JG	8,81
Barbera, S	7,85
Cabrales, A	7,79
Ganuza, JJ	7,08
Jackson, MO	6,89
Vorsatz, M	6,87
Corchon, L	6,27
Reynal-Querol, M	5,85
Klijn, F	5,07

TABLE 6

Top 10 Members of Galí Community

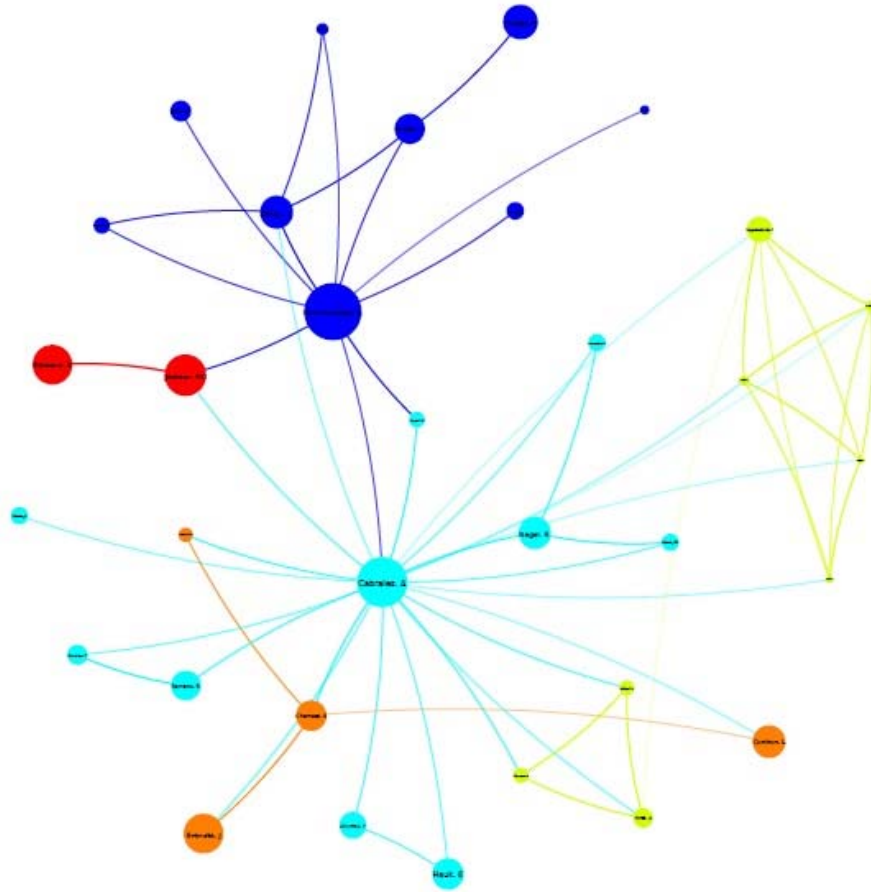| Researcher	Production
Galí, J	23,32
Domenech, R	5,06
Nakov, A	4,88
Arocena, P	4,45
De la Fuente, A	4,08
Andres, J	3,99
Nuno, G	3,88
Thomas, C	3,59
Castello-Climent, A	3,05
Ferri, J	2,77

FIGURE 9

Second Neighbours Network around Calvo-Armengol

FIGURE 10

Second Neighbours Network around Galí