Journal of Informetrics 11 (2017) 88–102

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Regular article

The specific shapes of gender imbalance in scientific authorships: A network approach

Tanya Araújo[a,b,*], Elsa Fontainha[a]

[a] ISEG (Lisbon School of Economics & Management), Universidade de Lisboa, Rua do Quelhas, 6, 1200-781 Lisboa, Portugal
[b] Research Unit on Complexity and Economics (UECE), Rua Miguel Lupi, 20, 1249-078 Lisboa, Portugal

ARTICLE INFO

Article history:
Received 13 August 2016
Accepted 12 November 2016
Available online 4 December 2016

Keywords:
Co-occurrence networks
Gender
Research collaboration
Interdisciplinarity
Bibliometrics
Minimum spanning tree

ABSTRACT

Gender differences in collaborative research have received little attention when compared with the growing importance that women hold in academia and research. Unsurprisingly, most of bibliometric databases have a strong lack of directly available information by gender. Although empirical-based network approaches are often used in the study of research collaboration, the studies about the influence of gender dissimilarities on the resulting topological outcomes are still scarce. Here, networks of scientific subjects are used to characterize patterns that might be associated to five categories of authorships which were built based on gender. We find enough evidence that gender imbalance in scientific authorships brings a peculiar trait to the networks induced from papers published in Web of Science (WoS) indexed journals of Economics over the period 2010–2015 and having at least one author affiliated to a Portuguese institution. Our results show the emergence of a specific pattern when the network of co-occurring subjects is induced from a set of papers exclusively authored by men. Such a male-exclusive authorship condition is found to be the solely responsible for the emergence of that particular shape in the network structure. This peculiar trait might facilitate future network analysis of research collaboration and interdisciplinarity.

1. Introduction

The handiness of powerful computational instruments and recent improvements in multidisciplinary methods are providing researchers an ever-greater opportunity to investigate societies in their complex nature (Banisch, Lima, & Araújo, 2012). Several research outcomes have been showing that men and women differ in characteristics that could be related to their collaboration patterns. Research collaboration is increasing in frequency and scope. It is driven, among other causes, by growing relationship across scientific disciplines, improvement of the efficiency in research resources in projects and development of information and communication technologies (Abramo, Cicero, & D'Angelo, 2015). The motivations (Beaver, 2001), strategies, patterns and impacts on scientific productivity in quantity and quality in research collaboration have received great scholarly attention (Börner, Dall'Asta, Ke, & Vespignani, 2005; Cainelli, Maggioni, Uberti, de, & Felice, 2015; Ductor, 2015). The patterns vary across space (Hoekman, Frenken, & Tijssen, 2010; Stefaniak, 2001), academic ranks (Abramo, D'Angelo, & Murgia, 2014), professional origins (Beaver & Rosen, 1978) and scientific disciplines (Tsai, Corley, & Bozeman, 2016).

* Corresponding author at: ISEG (Lisbon School of Economics & Management) Universidade de Lisboa, Rua do Quelhas, 6, 1200-781 Lisboa, Portugal.
E-mail address: tanya@iseg.ulisboa.pt (T. Araújo).

http://dx.doi.org/10.1016/j.joi.2016.11.002

Economic science makes connections with many other scientific disciplines, like Statistics or Social Sciences, like Sociology, History or Management (Krichel & Bakkalbasi, 2006; Pieters & Baumgartner, 2002). Economics shows a growing increase of co-authorship (Barnett, Ault, & Kaserman, 1988; Cainelli et al., 2015; McDowell & Melvin, 1983). On average, a researcher in Economics had less than one co-author in the 1970s, 1.24 co-authors in the 1980s and 1.67 in 1990s (Goyal, van der Leij, & Moraga-Gonzalez, 2006; Tsai et al., 2016).

Gender differences in collaborative research concerning motivations, strategies, patterns and impacts on science performance have received little attention, contrasting with the growing importance that women hold in academia and research. The literature shows mixed results about the gender differences concerning research collaboration strategies (McDowell & Melvin, 1983), impacts (Abramo et al., 2015; Frandsen, Jacobsen, Wallin, Brixen, & Ousager, 2015; Kyvik & Teigen, 1996; McDowell & Melvin, 1983; McDowell & Smith, 1992; Meng, 2016; Rorstad & Aksnes, 2015) and patterns (Abramo, D'Angelo, & Murgia, 2013; Barbezat, 2006; Boschini & Sjogren, 2007; Bozeman & Gaughan, 2011; Cottrell & Parpart, 2006; Kosmulski, 2015; Raasch, Lee, Spaeth, & Herstatt, 2013; Rhoten & Pfirman, 2007; Uhly, Visser, & Zippel, 2015).

Large bibliometric databases like Web of Science (Adriaanse & Rensleigh, 2013; Harzing & Alakangas, 2016; Sugimoto, Lariviere, Ni, Gingras, & Cronin, 2013) are the main sources used to bibliometric analysis. Bibliometric studies and survey analysis are the main methodologies to the study of research collaboration (Barabási et al., 2002). However, that bibliometric databases have a strong weakness concerning the study of the differences by gender; they do not include information separated by male–female and the way to overcome that weakness is to obtain the information from the first name (Naldi, Luzi, Valente, & Parenti, 2004) or the family name of the author (Kosmulski, 2015).

The present paper seeks to build upon the previous analysis about gender aspects in research collaboration which literature was recently surveyed in Abramo et al. (2013). Here, we intend to contribute to at least two points of the literature: the differences of research collaboration and interdisciplinary participation by gender. Focusing in Economics, a scientific subject strongly connected to other scientific domains (Pieters & Baumgartner, 2002) and constructing five categories of articles in a gender authorship perspective, this study addresses both issues: research collaboration and interdisciplinarity.

Applying a network approach and using as unit of analysis articles indexed in the Web of Science (WoS) this analysis maps the research collaboration by gender within dozen of scientific subjects, all associated with Economics.

Web of Science (WoS) and Scopus are the two major bibliographic databases (Wang & Waltman, 2016). WoS covers multiple types of scientific outputs. For example, the Social Sciences Citation Index (SSCI), for the period 2010-2015 and the Web of Science Subject 'Economics' includes 14 Document Types from which the Articles, Meeting Abstracts and Book Review correspond respectively to 76.7%, 13% and 5.9% of the total. It has been demonstrated that, in general, books are more important in Social Sciences and, in particular, in Art and Humanities than they are in Science (Chi, 2016). It is possible to extend bibliometric studies by using Library Catalog databases to focus on scholarly books in Economics, applying innovative methodologies (Torres-Salinas & Moed, 2009). Our research only includes articles published in English and in journals indexed to ISI-Thomson Reuters. While we recognize that it ignores books, non-English-language journals, local journals, monographs, confidential documents and 'grey' publications, there is a trend in Social Sciences towards publication in journals and away from monographs and similar documents (Norris & Oppenheim, 2007). In addition, concerning the non-English language journals included in the WoS Subject (WC) 'Economics' and based on a detailed analysis of each journal's language policy, Henshall (2012) shows that they account for a tiny proportion of the total. The latest updates in JCR reveal that this tiny proportion is further decreasing (Journal Citation Reports, years from 2010 to 2015). By using only one kind of scientific output (articles in English), our research follows previous literature that uses homogeneous information (Ruiz-Castillo & Waltman, 2015). The advantage of using articles published in ISI indexed journals instead of using other kinds of scientific outputs like books is that the selection process for journals included in Web of Science is public and relies on explicit publishing standards (Testa, 2016).

The choice of network approaches to study research collaboration in economics (Pieters & Baumgartner, 2002) has been extensively embraced. It often relies on the discovery of patterns of collaborations within researcher communities, aiming to find the influence of individual researchers in the networks using citation analysis. Beaver and Rosen (1978), in the first complete theory of scientific collaboration, list and discuss the causes for that collaboration.

Our unit of analysis is the article, not the journal. We define a multidisciplinary article as an article in the bibliographic database that includes Economics as WoS subject and at least one other WoS subject. (The list of co-occurrences with Economics in our database is presented in Table 1.) This multidisciplinary classification is completely independent from the thirteen WoS multidisciplinary categories (for example listed in Wang & Waltman, 2016:361, Table A1). The analysis of the accuracy and comparability across bibliometric databases of the Scientific Subject classifications is a relevant and crucial field of research (Leydesdorff & Bornmann, 2016; Wang & Waltman, 2016). We are aware of that important discussion, but in the current research we adopt the definition of a multidisciplinary article presented above assuming as given, and without discussing, the WoS Subjects (WCs).

They stress that it is necessary, when scientists deal with research questions, that cross disciplinary bounds. They also identify a large variation in collaboration by discipline, which is being further investigated in more recently published studies (Abramo et al., 2013; Bozeman & Gaughan, 2011).

Its well known that the adoption of a network approach allows the modeling of social structures from a bottom-up perspective, as resulting from the interaction (or likeness) of individual characteristics (Banisch et al., 2012). Moreover, as the individual characterization might be driven by multiple aggregate concerns, the network approaches allow for simultaneously considering that multiplicity of individual aspects and the consequences of the aggregate structures themselves on

Table 1
Secondary (or extra) subjects besides Economics.

Subject	Subject
1 Agricultural Economics	2 Area Studies
3 Business	4 Cultural Studies
5 Environmental Science	6 Education
7 Ecology	8 Finance
9 Geography	10 Health Policy
11 History of S.Sciences	12 Hospitality
13 Industrial Rel. &Labor	14 Interdisciplinary St.
15 International Relations	16 Leisure, Sport &Tourism
17 Management	18 Mathematics
19 Occupational Health	20 Operations Research
21 Planning &Development	22 Political Science
23 Science &Technology	24 Social Sciences
25 Sociology	26 Statistics &Probability
27 Transportation	28 Urban Studies
29 Engineering	

the emergence of collective patterns. Meanwhile, in the adoption of a network approach, one shall be aware that the choice of a given network representation is only one out of several other ways to look at a given set of elements. As connecting the elementary units of a system may be conceived in many different ways, that choice may depend strongly on the available empirical data and on the questions that a network analysis aims to address (Araújo & Banisch, 2016).

The main question addressed in this paper is whether some relevant characteristics of research collaboration would emerge in networks where subjects are linked whenever they co-occur in a common paper. We hypothesized that gender imbalance in authorship of papers might influence the shape of those networks, allowing to uncover patterns from gender differences. If it happens, the emerging patterns may help to understand important characteristics of research collaboration, of the relationship among subjects and its relation to gender.

The paper is organized as follows: next section presents the empirical data we work with and some preliminary statistical results. Section 3 describes the network approach and the results from its application. Section 4 concludes.

2. The data

The Web of Science (WoS) is one of the major bibliometric databases (together with Google Scholar and Scopus) and includes all scientific subjects. It comprises a total of 11,990 Journals (8778 from Science and 3212 from Social Sciences) (Harzing & Alakangas, 2016). The WoS classifies each journal in one or more subjects (or categories). Concerning the WoS Subject of Economics, it includes for each year in the period 2010–2015: 305 journals (2010), 321 journals (2011), 333 journals (2012–2014) and 345 journals (2015) (Journal Citation Reports, JCR). Taking as examples the journals Journal of Informetrics and Research Policy, the former is classified in “Computer Science, Interdisciplinary Applications” and “Information Science & Library Science”, while the latter is classified in “Management” and “Planning & Development”. From the original WoS database a selection of articles was carried on adopting as criteria: articles in English, published in WoS SSCI indexed journals over the period 2010–2015, having Economics as scientific subject according WoS Subject classification, and at least one author affiliated to a Portuguese institution.[1] Our motivation to focus on the field of Economics and on the papers whose authors are affiliated to Portuguese institutions is twofold:

1. Economic science makes connections with many other scientific subjects.
2. According OECD data, Portugal presents the highest percentage of women in research during the period of 2004–2012 (OECD, 2016).

Consequently, our approach is applied to a data set comprising 1138 papers published in 2010, 2011, 2012, 2013, 2014 and 2015 and having Economics as the main subject matter.

Besides Economics, each paper may have extra (or secondary) subjects. Table 1 presents the set of secondary (extra) subjects found in our data set. Each paper in the data set is coded by a string that informs about the presence of extra subjects. In the broader set of 1138 papers having Economics as the main subject matter, 29 different extra (or secondary) subjects were found. Some subjects in Table 1 were grouped merging WoS Subject Categories. Is the case of Engineering, Interdisciplinary Studies and Social Sciences. A full list of the WoS Subjects (for SSCI and SCI) is available at: http://admin-apps.webofknowledge.com/JCR/JCR.

Table 2
Exemplifying the representation of papers in the data set ($P_{(1138,5)}$).

id	#w	#m	1	2	3	4	5
0001	0	2	3	29	0	0	0
0002	1	0	1	0	0	0	0
...	...	...	...	...	...	...	...
1138	2	3	0	0	0	0	0

The structure presented in Table 2 exemplifies the way we represent the presence (and thus the co-occurrence) of subjects in each paper, it also shows the way we organize information on gender authorship.[2]

There, three papers are represented: the column *id* conveys the paper identification, the column #*w* stores the number of female authors, the column # *m* provides the number of male authors and the columns labeled 1, 2, . . ., 5 store the presence of extra subjects.

The examples in Table 2 inform that paper 0001 has two male authors and Business(3) and Engineering(29) as secondary (and co-occurring) subjects. It also informs that paper 0002 has just one female author and Agricultural Economics(1) as its single secondary subject. The paper 1138 has five authors: two female and three male authors and no extra subject.

As we aim to address interdisciplinarity issues, from the whole set of 1138 papers we select those the have at least one extra subject. They are 535 papers whose subjects are assembled in the set $P^0_{535,5}$. The superscript ^{0} identifies the subset of $P_{(1138,5)}$ that comprises all papers with at least one secondary subject. The first subscript (535) indicates the size of this data set while the second subscript (m) stands for the position of the extra subject in paper i with ($1 \leq m \leq 5$). There, each cell informs whether paper i has subject j ($p^0_{i,m} = j$) with $0 \leq j \leq 29$. Later in the paper, the set $P^0_{535,5}$ is used to construct the topological representation of the 29 subjects co-occurring with Economics in scientific publications.

2.1. Authorship categories

Besides the subject concerns and depending on the authorship characteristics, each paper belongs to at least one of the following (not mutually exclusive) categories. The definition of the five categories of authorship based on gender settles the basis for the identification of patterns of research collaboration and their relation to gender. The following list of categories is ranked in descending order of average percentage of female authors per article: 100, 51, 42, 20 and 0, respectively. The set papers belonging to the authorship categories are labeled $P^1_{(57,5)}$, $P^2_{(266,5)}$, $P^3_{(209,5)}$, $P^4_{(478,5)}$ and $P^5_{(269,5)}$, respectively.

1. $P^1_{(57,5)}$: all authors are women (W.Exc)
2. $P^2_{(266,5)}$: authors include at least one woman (W.Inc)
3. $P^3_{(209,5)}$: authors include both women and men (W&M)
4. $P^4_{(478,5)}$: authors include at least one man (M.Inc)
5. $P^5_{(269,5)}$: all authors are men (M.Exc)

Considering the articles in each category, some statistical values are computed:

- the number of articles (Size)
- the average number of authors per article (<Author>)
- the average percentage of female authors per article (% female)
- the number of articles with a single author (#Single)
- the average number of subjects by article (<Subject>)
- the number of articles with at least one extra subject (XSubject)

2.2. Overview of the data set

Table 3 shows the overall statistics for the 1138 papers from 2010 to 2015 in Economics, according to the five authorship categories above presented. While the columns correspond to the authorship categories, the rows in Table 3 provide the values obtained for the statistical indicators above described. The sum of the articles in three categories (W.Exc, M.Exc and W&M, respectively 105, 613 and 420) corresponds to the total of the sample (1138 articles).

[1] The advanced search used was:(CU = Portugal AND WC = Economics AND (PY = 2010 OR PY = 2011 OR PY = 2012 OR PY = 2013 OR PY = 2014 OR PY = 2015)) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article) Timespan: 2010–2015. Indexes: SSCI.

[2] The gender of the authors was identified by the first given name, because in Portuguese, the first given name defines the gender without any ambiguity. When the authors did not have Portuguese given names, the identification was made by visiting the institutional web pages of each of the authors.

Table 3
Overall Statistics for 2010–2015 papers in Economics.

Authorship	All	W.Exc	W.Inc	W&M	M.Inc	M.Exc
Category	0	1	2	3	4	5
Size	1138	105	525	420	1033	613
<Author>	2.4	1.8	2.7	3	2.5	2.1
% female	25	100	51	42	20	0
Single	210	46				164
<Subject>	2	1.75	1.9	1	2	2.1
XSubject	535	57	266	209	478	269

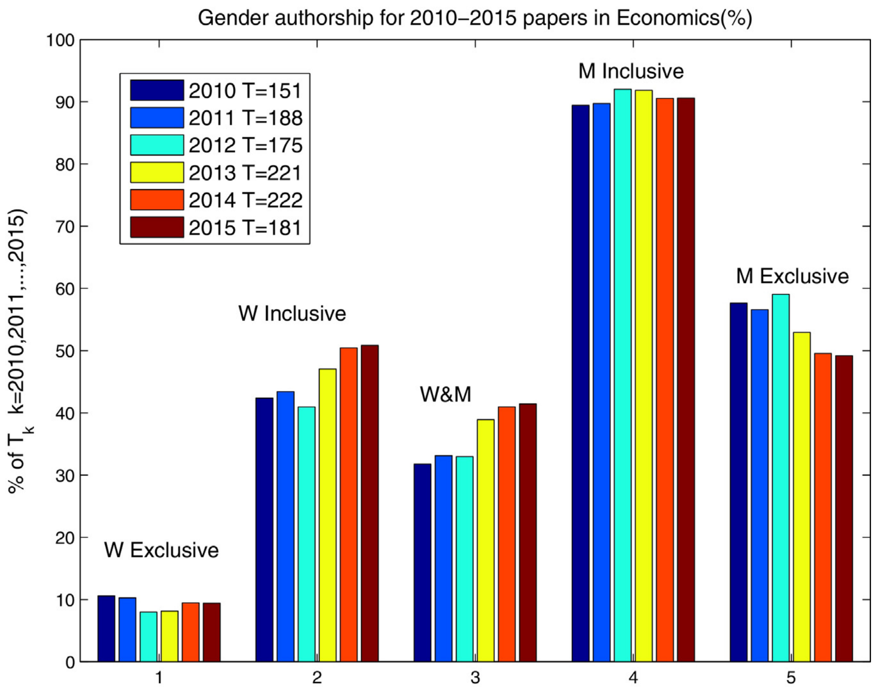

Fig. 1. The distribution of the percentage (% in each year) of papers in each category for papers in Economics.

The results in Table 3 seem to contradict the hypothesis that women have more propensity to interdisciplinary research collaboration, because the category man exclusive (M.Exc) is the one which has the higher average number of subjects. The average number of authors is higher in the mixed category W&M but the woman inclusive (W.Inc) is the category with second highest number of authors (the size of co-authorship). These results apparently converge to the hypothesis that women prefer to work in teams. However, this hypothesis is not confirmed by the average number of authors of the papers in the woman exclusive category (W.Exc), being the smallest value in the <Author> row, it indicates that when papers are exclusively authored by women, the working teams tend to be smaller than any of those that also include men. Looking at the number of papers authored by a single individual (210 papers), 22% and 78% are the respective percentages of female and male authorships. A similar proportionality characterizes the percentages of woman exclusive and man exclusive authorships (W.Exc and M.Exc) in the total amount of papers in these two exclusive categories, they are 25% (105 papers) and 75% (316 papers), respectively.

Figs. 1 and 2 show the dynamic of the five categories of authorship across time (2010–2015). Fig. 1 displays the distributions of the percentage of all articles ($N = 1138$) in each authorship category. The plots in Fig. 2 show: (a) the distributions of the percentage (values per year) of articles in Economics with at least one secondary subject ($N = 535$); and the distribution of all articles in Economics with a single author ($N = 210$).

The distributions in Figs. 1 and 2 are quite similar meaning that constraining our sample to the papers with at least one extra subject does not introduce any bias, the only (and unimportant) exception regards the man inclusive category (M.Inc) in the first two years. The same would apply to the distributions presented in Fig. 2 if the year of 2014 was excluded. In 2014 the proportions of gender-based single authorship shows a different balance between male and female authorships (moving from 32% and 5% to 25% and 10%, respectively). As presented in the last rows of Table 3, the set of papers presenting a least one extra subject comprise 535 papers and the average number of extra subjects by paper in this set is 2.

In general, the Figs. 1 and 2(a) reveal that there is an increasing trend in all co-authorship categories, and a relative decrease in M. Exclusive category. Note that, due to indexing delay, it is likely that the results for recent years are incomplete.

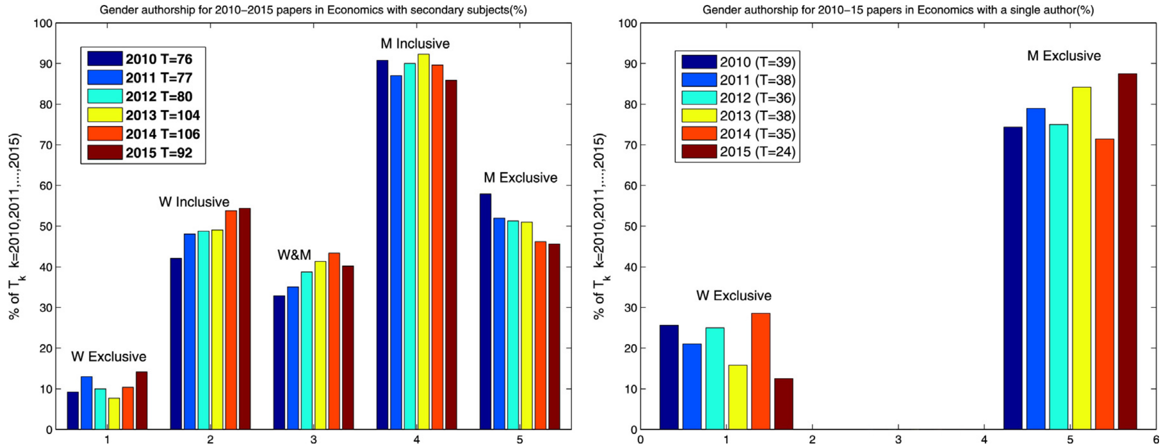

Fig. 2. The distributions of the percentage (% in each year) of papers (a) in Economics with at least one secondary subject and (b) in Economics with a single author.

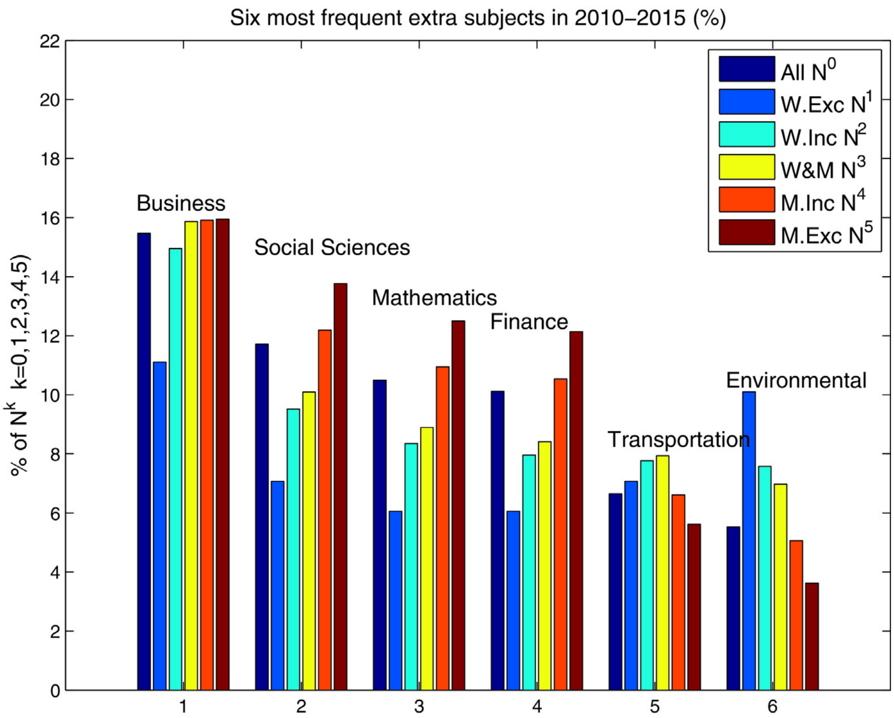

Fig. 3. The distribution of the frequencies of the six most frequent extra subjects in each authorship category.

Fig. 3 shows the distributions of the relative frequencies (%) of the six most frequent extra subjects in each authorship category. Fig. 4 shows the distributions of the relative frequencies (%) of the 7th to the 12th most frequent extra subjects in each authorship category.

These distributions show that the exclusive categories W.Exc (dark blue) and M.Exc (red) display the greater fluctuations along the different subjects. These fluctuations increase from the 5th most frequent subject (Transportation) until the 10th (Political Sciences). The larger imbalance between the relative frequencies of the exclusive categories W.Exc and M.Exc relies on the subjects Environmental Sciences, Management and Political Sciences. When compared with the high homogeneous distribution that characterizes Business, the relative frequencies of Environmental Sciences, Management and Political Sciences increase in the woman exclusive category (W.Exc) in the same proportion they decrease in the man exclusive (M.Exc) one. These very first results indicate that the subjects Environmental Sciences, Management and Political Sciences are more likely to co-occur in female-dominated papers in Economics.

In summary, when considering papers published in WoS indexed journals over the period 2010–2015 in the scientific domain of Economics and whose authors are affiliated to a Portuguese institution, our results suggest that:

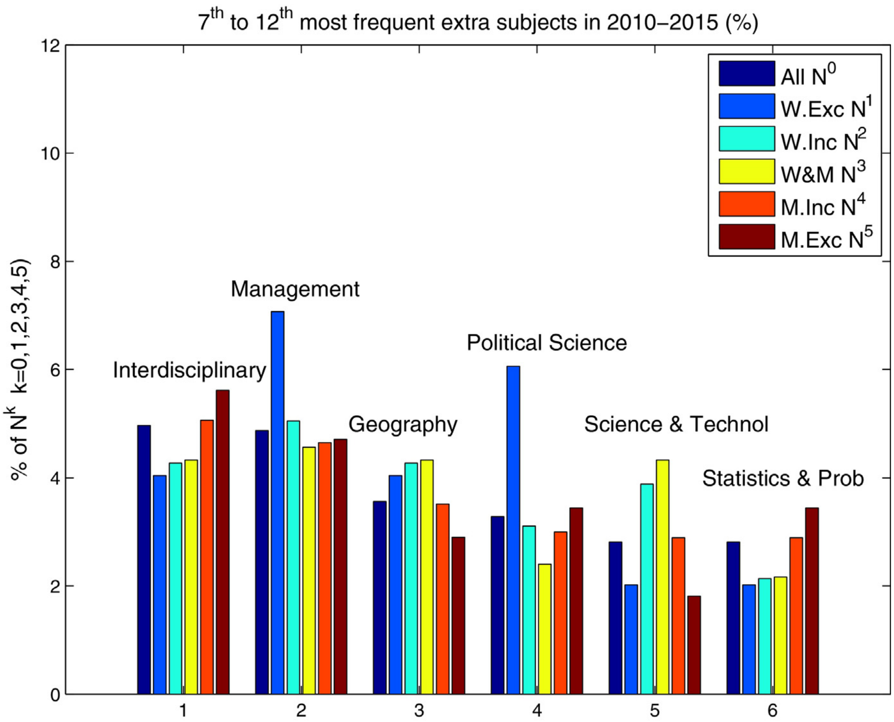

Fig. 4. The distribution of the frequencies (%) of the seventh to the twelfth most frequent extra subjects in each authorship category.

1. men have more propensity to interdisciplinary research collaboration, since the man exclusive category (M.Exc) has the highest average number of subjects
2. the woman inclusive (W.Inc) is the category with the second highest average number of authors. These results apparently converge to the hypothesis that women prefer to work in teams but
3. when papers are exclusively authored by women (W.Exc), the working teams tend to be smaller than any of those that also include men
4. academic women compared with their male counterparts reveal preference for the subjects Environmental Sciences, Management and Political Sciences
5. conversely, the subjects Social Sciences, Mathematics and Finance display higher frequencies in papers either inclusively (M.Inc) or exclusively authored by men (M.Exc)

In the next section, a network approach is applied to combine the gender authorship perspective with the analysis of interdisciplinarity. To this end, the categories of articles are used to construct the topological representation of the 29 subjects (Table 1) co-occurring with Economics in scientific publications.

3. Network induction

Network induction makes reference to the method by which networks are created on the basis of a certain data set or system (Araújo & Banisch, 2016). As earlier mentioned, network approaches are quite common in the analysis of systems where a network representation is the most intuitive. As connecting the elementary units of a system may be conceived in many different ways, that choice may depend strongly on the available empirical data and on the questions that a network analysis aims to address. Here, six bipartite networks are induced from the subsets of papers defined by the authorship categories presented in Section 2.1.

The frequency of co-occurrence of each pair of subjects defines the existence of every link in the networks by authorship category. They are weighted graphs since the weight of each link corresponds to the frequency of co-occurrence of the linked pair of subjects. In the next section, those weighted networks are further analyzed through the construction of their corresponding minimum spanning trees (MST). In so doing, we are able to emphasize the main topological patterns that emerge from each network representation and to discuss their interpretation and relation to gender.

3.1. Bipartite graphs

A bipartite network N consists of two partitions of nodes V and W, such that edges connect nodes from different partitions, but never those in the same partition. A one-mode projection of such a bipartite network onto V is a network consisting of the nodes in V; two nodes v and $v\prime$ are connected in the one-mode projection, if and only if there exist a node $w \in W$ such that (v, w) and $(v\prime, w)$ are edges in the corresponding bipartite network (N). In the following, we explore six bipartite networks and their corresponding one-mode projections.

3.2. Connecting subjects

Each bipartite network by authorship category consists of the following partitions:

- the set S of 29 subjects presented in Table 1 and
- one set of papers (P^k) by authorship category ($k = \{0, 1, 2, 3, 4, 5\}$) presented in Section 2.1.

In the each network (N^k), two subjects are linked if and only if they co-occur in at least one paper of P^k, having each paper at most five subjects. Therefore, the links in each network (N^k) are weighted by the number of coincident papers a pair of subjects share in P^k. Consequently, every link $L^k_{(i,j)}$ in N^k takes value in the set { 0,1,2,. . ., size(P^k)}.

As an example and considering that in $P^0_{(535,5)}$(the authorship category comprising all papers with at least one secondary subject) there are just three papers where the subjects Agricultural Economics and Finance co-occur yields $L^0_{(1,8)} = 3$. Another example is $L^0_{(1,3)} = 1$ due to the mutual single co-occurrence of Agricultural Economics and Business in $P^0_{(535,5)}$. Among the many examples of missing links there are the cases of and Education and Finance($L^0_{(6,8)} = 0$) since these two subject do not co-occur in any paper of $P^0_{(535,5)}$.

Having induced the networks (N^k) for each authorship category, we are able to have a complete representation of the relationship among the subjects co-occurring in each authorship category defined based on gender. However, it so happens that neither the densely-connected nature nor the existence of disconnected components of these networks helps to find out whether there is a dominant pattern in the structure of subjects. The large number of links make the extraction of the truly relevant connections forming the network a challenging problem. One first step in the direction of extracting relevant information from the networks may be targeted at obtaining the corresponding MST (Araújo & Ferreira, 2016; Araújo & Louç a, 2007; Araújo & Spelta, 2012; Vilela Mendes, Araújo, & Louçã, 2003).

3.2.1. From complete networks to minimum spanning trees

In the construction of a MST by the *nearest neighbor* method, one defines the subjects (in Table 1) as the nodes (n^k_i) of a weighted and connected[3] network (N^k) where the distance d^k_{ij} between each pair of subjects i and j corresponds to the inverse of weight of the link ($d^k_{ij} = 1/L^k_{ij}$) between i and j.

From the nxn distance matrix $D^k_{i,j}$, a hierarchical clustering is performed using the *nearest neighbor* method. Initially n clusters corresponding to the n subjects are considered. Then, at each step, two clusters c_i and c_j are clumped into a single cluster if

$$d^k\{c_i, c_j\} = \min\{d^k\{c_i, c_j\}\}$$

with the distance between clusters being defined by

$$d^k\{c_i, c_j\} = \min\{d^k_{pq}\} \quad \text{with}\, p \in c_i and q \in c_j$$

This process is continued until there is a single cluster. This clustering process is also known as the *single link method*, being the method by which one obtains the MST of a graph (Araújo & Vilela Mendes, 2000).

In a connected graph, the MST is a tree of $n - 1$ edges that minimizes the sum of the edge distances. In a network with n nodes, the hierarchical clustering process takes $n - 1$ steps to be completed, and uses, at each step, a particular distance $d^k_{i,j} \in D^k$ to clump two clusters into a single one.

Let $C = \{d_q\}, q = 1, \ldots, N - 1$, be the set of distances $d^k_{i,j} \in D^k$ used at each step of the clustering, and $thr = \max\{d_q\}$. It follows that $thr = d^k_{N-1}$.

The result of the hierarchical clustering process leading to the MST is usually described by means of a dendrogram. During this process, a unique color is assigned to each group of nodes within the dendrogram whose linkage is less than T times the value of the threshold distance (thr). In the dendrogram presented here, T is set to 1.2.

[3] The hierarchical clustering process considers just the largest connected component of each network (N^k). Therefore, depending on the authorship category (k) the resulting MST has different size, as indicated in the first row of Table 4.

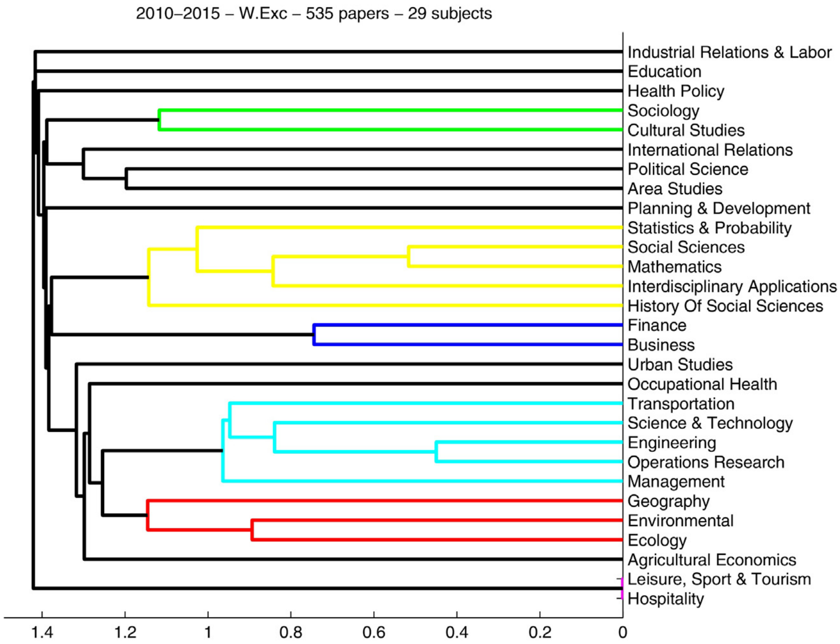

Fig. 5. The dendrogram shows the hierarchical clustering process applied to N^0.

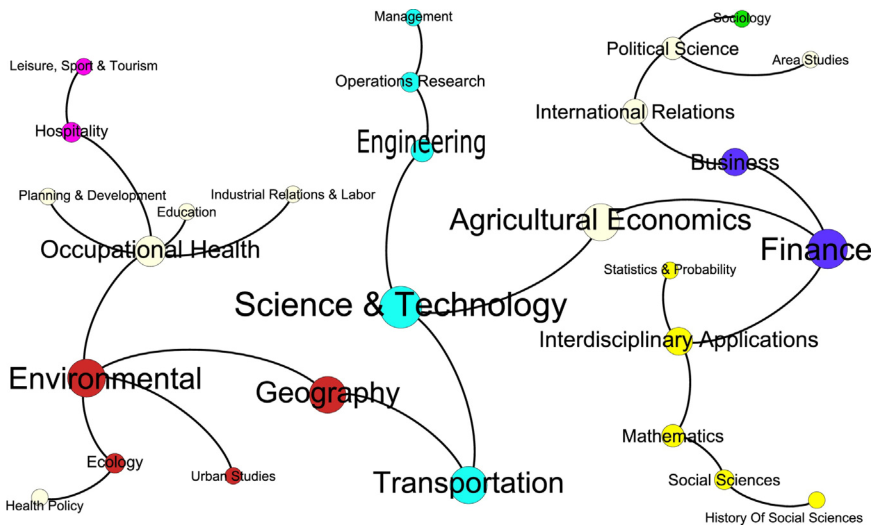

Fig. 6. The MST of N^0, which comprises all papers with at least one secondary subject.

Six clusters can be observed in the dendrogram of N^0(the network of the authorship category comprising all papers with at least one secondary subject) as Fig. 5 shows. The colors assigned to these clusters will be hereafter used in the identification of the same partitions of subjects whenever represented in a MST.

The dendrogram in Fig. 5 shows that the subjects Hospitality and Leisure, Sports & Tourism are the first to be clumped since their occurrences are perfectly correlated in $P^0_{(535,5)}$. On the other hand, the papers on these two subjects remain almost isolated from any other subject matter in the overall set of papers being considered. The next cluster being defined comprises the subjects Business and Finance (colored blue). Being followed by the large cluster including Mathematics, Statistics, Social Sciences and Interdisciplinary Sciences (yellow). Another early defined cluster clumps together Transportation, Operational Research, Engineering and Science & Technology (turquoise). Further analyzing a dendrogram by its corresponding MST allows for observing the extent to which clusters give place to branches on the tree and whether different motifs emerge from the clusters positioning on the trees.

Table 4
Topological coefficients computed from the MST of each authorship category.

Authorship category MST	1 W.Exc	2 W.Inc	3 W&M	4 M.Inc	5 M.Exc
N	25	28	27	29	27
d	13	12	11	12	17
l	11	11	12	13	10
$\frac{d}{N-1}$	0.54	0.44	0.44	0.43	0.65
% female authors	100	51	42	23	0

Fig. 6 shows the representation of the corresponding MST. It is worth noting that closeness on the MST depends on the connection strength (the weight of the links) in N^0, meaning that when two subjects co-occur in many papers of $P^0_{(535,5)}$ (being therefore strongly connected) they occupy close positions on this tree.

While the dendrograms provide information on the distances at which the subjects are clumped into clusters, their corresponding minimum spanning trees allow for the identification of at least four important aspects that are not directly stated in the dendrograms.

1. Branches: the way nodes organize themselves in different ramifications of the tree
2. Motifs: the prevalence of *star* motifs and/or *path* motifs in the tree
3. Connectivity: highly connected and weakly connected nodes
4. Centrality: the nodes occupying highly central positions and, conversely, those occupying the leafs of the tree

The observation of the MST presented in Fig. 6 suggests that besides a "core" cluster, there are at least three important branches separating "classical" subjects; "technological" subjects and "environment-related" subjects. Fig. 6 also shows that there are two highly connected nodes: Environmental Sciences and Cultural Studies.

In what concerns centrality, the subjects Political Sciences and Science & Technology occupy positions of great centrality on the tree. A distinct situation characterizes Education and Industrial Relations & Labor which occupy leaf positions on the MST. These two subjects, together with the cluster that joins Hospitality and Leisure, Sports & Tourism are the last ones to be connected in the hierarchical clustering process, as the dendrogram of Fig. 5 shows.

3.2.2. The minimum spanning trees by authorship category

Since we hypothesized that specific characteristics could come out and shape the structures of the networks of subjects and that these characteristics may be associated to some ordering emerging from gender, here we consider the subsets of papers defined by the authorship categories $P^1_{(57,5)}$, $P^2_{(266,5)}$, $P^3_{(209,5)}$, $P^4_{(478,5)}$ and $P^5_{(269,5)}$. In applying the hierarchical clustering process to each subset provides the following MSTs. They are ranked in descending of average percentage of female authors per article (as in Section 2.1).

1. All authors are women (W.Exc-MST)
2. Authors include at least one woman (W.Inc-MST)
3. Authors include both women and men (W&M-MST)
4. Authors include at least one man (M.Inc-MST)
5. All authors are men (M.Exc-MST)

Obtaining the MST of a given network implies that the network is connected. Therefore, the application of the hierarchical clustering process to each network (N^k) by authorship category considers just the largest connected component of each network. Thereafter, depending on the authorship category (k) the resulting MSTs have different sizes, which are indicated in the first row of Table 4 (N^0).

Figs. 7 and 8 present the minimum spanning trees of the gender exclusive authorship categories (W.Exc-MST and M.Exc-MST), being the nodes colored according to the partitions of subjects as defined in the dendrogram presented in Fig. 5.

These networks are quite similar in the way nodes organize themselves in different branches (clusters) on the tree. However, there is an important difference concerning the centrality of certain nodes and the positioning of the main branches on the trees.

When centrality matters, Management occupies a central position in the woman exclusive (W.Exc-MST in Fig. 7) but looses centrality in the man exclusive one (M.Exc-MST in Fig. 8). The positioning of the "core", "classical" and "technological" branches suffer important changes when compared to their situation in the global MST (N^0 in Fig. 6). While the "core" and the "classical" branches remain linked in both the female and the man exclusive, the "technological" and the "core" branches, that in the global MST were linked through the Agricultural node are far away in the man exclusive MST (M.Exc-MST). The fact that they occupy close positions on the woman exclusive MST (W.Exc-MST) is probably associated to the greater centrality of the subject Management in this tree.

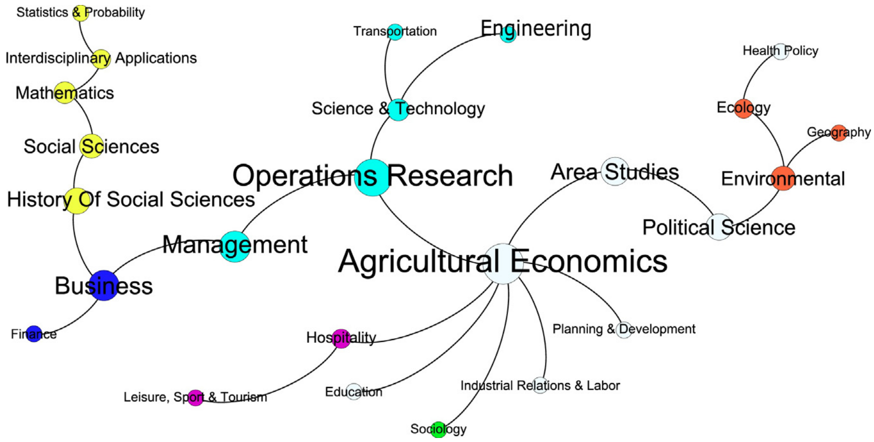

Fig. 7. The MST of the woman exclusive category (W.Exc-MST).

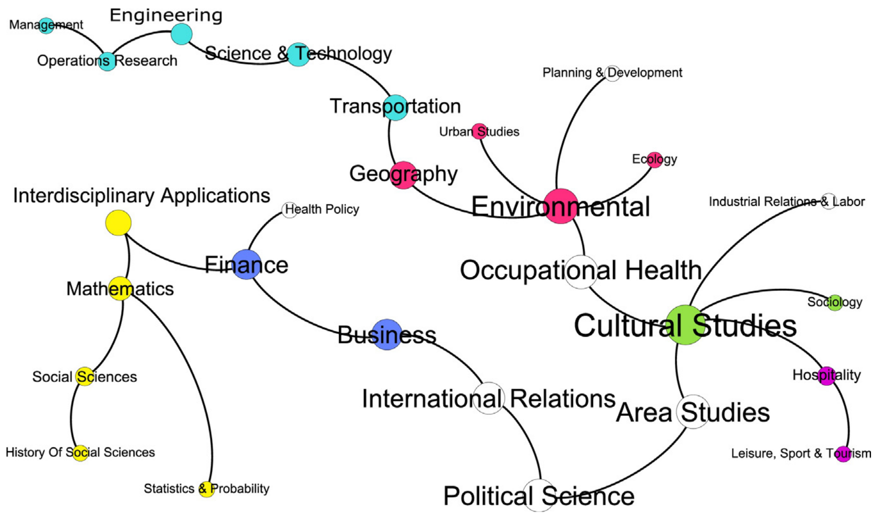

Fig. 8. The MST of the man exclusive category (M.Exc-MST).

The increase of centrality of the subject Management in the woman exclusive MST together with the presence of the subject Agricultural Economics has an important bearing on that tree (W.Exc-MST), showing that when papers authorship includes just women, the larger distances between subjects in the network tend to be reduced due to an important increase in the relative number of papers having Management as a secondary subject.

By constructing the MST from the largest connected component of a network allows for the identification of branches and motifs. Moreover, it allows for highlighting assortativity mixing characteristics of M.Exc-MST and W.Exc-MST for the node degree. Assortativity mixing of a network is a measure of the extent to which a node with a higher value for a particular node level measure is connected to another node that also has a higher value for the node level measure (Newman, 1999). Here, the node level measure is the degree (k). The observation of the networks in Figs. 7 and 8, shows that assortativity mixing (for k) is stronger in the Woman Exclusive network. In the W.Exc-MST, three out of five nodes with the highest degrees happen to be connected while in the M.Exc-MST (and N0-MST in Fig. 6), the most connected nodes are not linked to each other. This result is in line with our previous finding that when papers authorship excludes women, the larger distances between subjects in the network tend to be even larger.

3.2.3. Tree motifs

The adoption of a network approach provides well-known notions of graph theory to fully characterize the structure of the networks. Here, and since our analysis relies on the minimum spanning trees, we concentrate on the calculation of just two topological coefficients, both measured at the network level.

The first one is the number of leafs (l) in the MST, i.e., the number of nodes with degree one. The second coefficient is the MST diameter (d), measuring the shortest distance between the two most distant nodes on the tree. The choice of these coefficients allows to characterize tree motifs with different shapes: from a pure *star* to a pure *path* motif. Fig. 9 shows

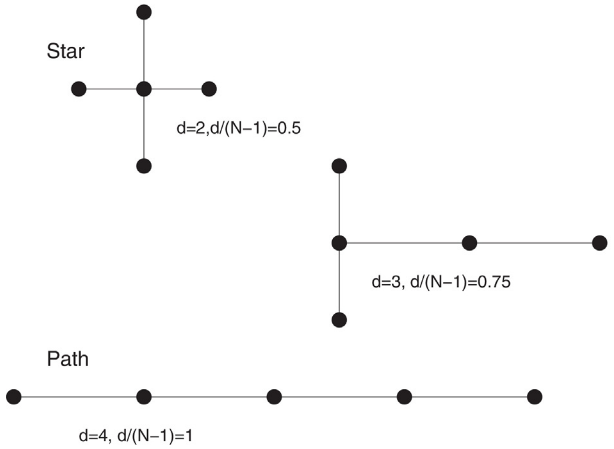

Fig. 9. Examples of different motifs of a tree with five nodes: from a Star to a Path motif.

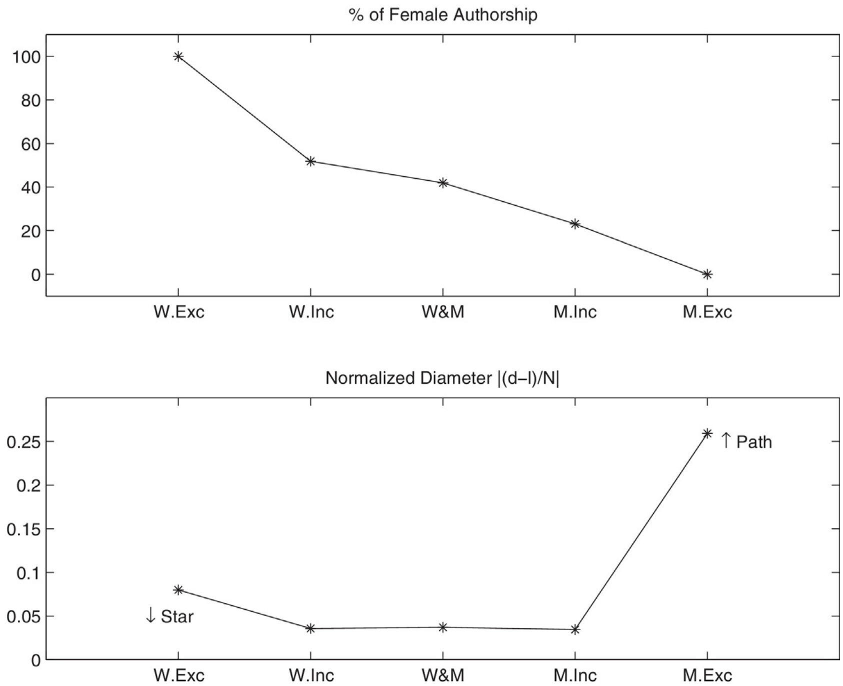

Fig. 10. The % of female authorship along with the different MST categories and the corresponding evolution of $|d-l|/N$.

examples of different motifs occurring on a tree of just five nodes ($N=5$) and the values of each corresponding diameter (d) and the normalized diameter $d/N-1$.

It so happens that when the number of nodes of the tree is greater than 2, and depending on the motif that the MST approaches, its diameter ranges in between 2 and $N-1$ ($2 \le d \le N-1$). The closer is $d/N-1$to 1, the smaller is the similarity of the MST to a *star* motif. Moreover, the number of leafs ranges in between exactly the same values but in the opposite direction, the closer l is to 1, the smaller is the similarity of the MST to a *path* motif.

Table 4 shows the values of N, d, l and $d/N-1$ computed for the five trees by the authorship category. The first row in Table 4 displays the size of each MST, i.e., the number of nodes in each MST. The last row shows the values obtained for the normalized diameter $d/N-1$, which are limited between $2/N-1$(*star*) and 1 (*path*).

Although the five networks have similar sizes, there is a remarkable difference in the values obtained for the man exclusive tree (M.Exc-MST). When women are excluded, the network of subjects displays an much higher diameter (d), showing large distances among subjects are enlarged. The also important decrease in the number of leafs (l) indicates that this network develops a entirely different structure when compared with the other MSTs by authorship category.

The plots in Fig. 10 show the number of leafs (l), the diameter (d) and the normalized diameter $|d-l|/N$ across the different categories of authorship. As, depending on the specific tree motif, the values of d and l move in opposite directions, in computing the absolute value of the difference $d-l$ relative to N helps to emphasize the distinguish structure of the MST that characterizes the man exclusive network (M.Exc-MST).

In the broader set of papers published in WoS indexed journals over the period 2010-2015 in the scientific domain of Economics and having at least one author affiliated to a Portuguese institution, as the percentage of female authorship decreases, the MST obtained from the corresponding network of subjects moves from a star configuration to a path motif. In so doing, the larger distances between subjects are enlarged and the number of poorly connected subjects increases. If, conversely, the network of subjects has a high percentage of female authorship, the corresponding MST approaches a star motif, the number of leafs is enlarged and the corresponding diameter decreases.[4]

4. Conclusion

There are many ways to link the elementary units of system in order to induce a network. Choosing the most suitable way depends strongly on the available empirical data and on the research questions that a network analysis aims to address. Regarding available empirical data, most of bibliometric databases have a strong weakness concerning the study of the differences by gender. In what concerns research questions, gender differences in collaborative research and interdisciplinarity in scientific outputs have received little attention when compared with the growing importance that women hold in academia and research.

From the set of papers published in WoS indexed journals over the period 2010–2015 in the scientific domain of Economics and having at least one author affiliated to a Portuguese institution, our results apparently converge to the hypothesis that women prefer to work in teams. However, they also indicate that when papers are exclusively authored by women, the working teams tend to be smaller than any of those that also include men. These results converge to the mixed results reported in the literature, where different units of analysis, measures, methods and samples were adopted (Abramo et al., 2015; McDowell & Melvin, 1983).

Regarding interdisciplinarity, our findings seem to contradict the hypothesis that women have more propensity to interdisciplinary research collaboration (Abramo et al., 2015). Moreover, we found that academic women in Economics compared with their male counterparts reveal preference for the subjects Environmental Sciences, Management and Political Sciences and that, conversely, the subjects Social Sciences, Mathematics and Finance display higher frequencies in papers either inclusively or exclusively authored by men.

Our main contribution relies in the adoption of a network approach allowing to uncover the emergence of a specific pattern when the network of scientific subjects is induced from a set of papers exclusively authored by men. Such a male exclusive authorship condition is found to be the solely responsible for the emergence of that specific shape in the structure of the network.

Moving away from a *star* motif together with the loss of centrality of the subject Management have an important bearing on the structure of the male exclusive authorship network: when papers authorship includes just men, the larger distances between subjects in the network become even larger and this is mainly due to a decrease in the relative number of papers having Management as a secondary subject. We find enough evidence that gender imbalance in scientific authorships brings a peculiar trait to the networks of subjects. Such a peculiar trait might facilitate future network analysis of research collaboration and interdisciplinarity.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgements

Financial support by FCT (Fundação para a Ciência e a Tecnologia), Portugal is gratefully acknowledged. This article is part of the Strategic Project: UID/ECO/00436/2013. The authors thank Marilei Kroetz for her work in the gender identification process.

Author contributions

Conceived and designed the analysis: Tanya Araújo.
Collected the data: Elsa Fontainha.
Contributed data or analysis tools: Tanya Araújo, Elsa Fontainha.
Performed the analysis: Tanya Araújo.
Wrote the paper: Tanya Araújo, Elsa Fontainha.

[4] The woman exclusive authorship (W.Exc-MST) shows a small deviation in the opposite direction. However, the fact that this network was induced from a small number of papers (57) might introduce some bias in its shape.

References

Abramo, G., D'Angelo, C. A., & Murgia, G. (2013). Gender differences in research collaboration. *Journal of Informetrics*, *7*, 811–822. http://dx.doi.org/10.1016/j.joi.2013.07.002

Abramo, G., D'Angelo, C. A., & Murgia, G. (2014). Variation in research collaboration patterns across academic ranks. *Scientometrics*, *98*, 2275–2294. http://dx.doi.org/10.1007/s11192-013-1185-3

Abramo, G., Cicero, T., & D'Angelo, C. A. (2015). Should the research performance of scientists be distinguished by gender? *Journal of Informetrics*, *9*, 25–38. http://dx.doi.org/10.1016/j.joi.2014.11.002

Adriaanse, L. S., & Rensleigh, C. (2013). Web of Science, Scopus and Google Scholar. A content comprehensiveness comparison. *Electronic Library*, *31*, 727–744. http://dx.doi.org/10.1108/EL-12-2011-0174

Araújo, T., & Banisch, S. (2016). Multidimensional analysis of linguistic networks. In *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*. pp. 107–131. Berlin/Heidelberg: Springer. http://dx.doi.org/10.1007/978-3-662-47238-5

Araújo, T., & Ferreira, M. E. (2016). The topology of African exports: Emerging patterns on spanning trees. *Physica A*, *462*, 962–976. http://dx.doi.org/10.1016/j.physa.2016.06.44

Araújo, T., & Louç a, F. (2007). The geometry of crashes – A measure of the dynamics of stock market crises. *Quantitative Finance*, *7*, 63–74. http://dx.doi.org/10.1080/14697680601019530

Araújo, T., & Spelta, A. (2012). The topology of cross-border exposures: Beyond the minimal spanning tree approach. *Physica A*, *391*, 5572–5583. http://dx.doi.org/10.1016/j.physa.2012.05.071

Araújo, T., & Vilela Mendes, R. (2000). Function and form in networks of interacting agents. *Complex Systems*, *12*, 357–373. http://arXiv:nlin/0009018v1

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*, 57–67. http://dx.doi.org/10.1002/cplx.20078

Banisch, S., Lima, R., & Araújo, T. (2012). Agent based models and opinion dynamics as Markov chains. *Social Networks*, *34*, 549–561. http://dx.doi.org/10.1016/j.socnet.2012.06.001

Barabási, A. L., Jeong, H., Ravasz, R., Néda, Z., Vicsek, T., & Schubert, A. (2002). On the topology of the scientific collaboration networks. *Physica A*, *311*, 590–614. http://dx.doi.org/10.1016/S0378-4371(02)00736-7, 4–6

Barbezat, D. A. (2006). Gender differences in research patterns among PhD economists. *Journal of Economic Education*, *37*, 359–374. http://dx.doi.org/10.3200/JECE.37.3.359-375

Barnett, A. H., Ault, R. W., & Kaserman, D. L. (1988). The rising incidence of co-authorship in economics-further evidence. *Review of Economics and Statistics*, *70*, 539–543. http://dx.doi.org/10.2307/1926798

Beaver, D. D., & Rosen, R. (1978). Studies in scientific collaboration: Part I. The professional origins of scientific co-authorship. *Scientometrics*, *1*, 65–84. http://dx.doi.org/10.1007/BF02016840

Beaver, D. D. (2001). Reflections on scientific collaboration (and its study): Past, present, and future. *Scientometrics*, *52*, 365–377. http://dx.doi.org/10.1023/A:1014254214337

Boschini, A., & Sjogren, A. (2007). Is team formation gender neutral? Evidence from coauthorship patterns. *Journal of Labor Economics*, *25*, 325–365. http://dx.doi.org/10.1086/510764

Bozeman, B., & Gaughan, M. (2011). How do men and women differ in research collaborations? An analysis of the collaborative motives and strategies of academic researchers. *Research Policy*, *40*, 1393–1402. http://dx.doi.org/10.1016/j.respol.2011.07.002

Cainelli, G., Maggioni, M. A., Uberti, T. E., de, & Felice, A. (2015). The strength of strong ties: How co-authorship affect productivity of academic economists? *Scientometrics*, *102*, 673–699. http://dx.doi.org/10.1007/s11192-014-1421-5

Chi, P. S. (2016). Differing disciplinary citation concentration patterns of book and journal literature? *Journal of Informetrics*, *10*(3), 814–829. http://dx.doi.org/10.1016/j.joi.2016.05.005

Cottrell, B., & Parpart, J. L. (2006). Academic-community collaboration, gender research, and development: Pitfalls and possibilities. *Development in Practice*, *1*, 15–26. http://dx.doi.org/10.1080/09614520500450768

Ductor, L. (2015). Does co-authorship lead to higher academic productivity? *Oxford Bulletin of Economics and Statistics*, *77*, 385–407. http://dx.doi.org/10.1111/obes.12070

Frandsen, T. F., Jacobsen, R. H., Wallin, J. A., Brixen, K., & Ousager, J. (2015). Gender differences in scientific performance: A bibliometric matching analysis of Danish health sciences Graduates. *Journal of Informetrics*, *9*, 1007–1017. http://dx.doi.org/10.1016/j.joi.2015.09.006

Goyal, S., van der Leij, M. J., & Moraga-Gonzalez, J. L. (2006). Economics: An emerging small world. *Journal of Political Economy*, *114*, 403–412. http://dx.doi.org/10.1086/500990

Harzing, A.-W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, *106*, 787–804. http://dx.doi.org/10.1007/s11192-015-1798-9

Henshall, A. C. (2012). English as an international language and language policies in economics journals. *Revista Anglo Saxonica*, *III*(4), 131–164. http://www.ulices.org/images/site/publicacoes/anglo-saxonica/

Hoekman, J., Frenken, K., & Tijssen, R. J. W. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy*, *39*, 662–673. http://dx.doi.org/10.1016/j.respol.2010.01.012

Kosmulski, M. (2015). Gender disparity in Polish science by year (1975–2014) and by discipline. *Journal of Informetrics*, *9*, 658–666. http://dx.doi.org/10.1016/j.joi.2015.07.010

Krichel, T., & Bakkalbasi, N. (2006). A social network analysis of research collaboration in the economics community. *Journal of Information Management and Scientometrics*, *3*, 1–12.

Kyvik, S., & Teigen, M. (1996). Child care, research collaboration, and gender differences in scientific productivity. *Science Technology & Human Values*, *21*, 54–71. http://dx.doi.org/10.1177/016224399602100103

Leydesdorff, L., & Bornmann, L. (2016). The operationalization of fields as WoS subject categories (WCs) in evaluative bibliometrics: The cases of Library and Information Science and Science & Technology Studies. *Journal of the Association for Information Science and Technology*, *67*(3), 707–714. http://dx.doi.org/10.1002/asi.23408

McDowell, J. M., & Melvin, M. (1983). The determinants of co-authorship – An analysis of the economics literature. *Review of Economics and Statistics*, *65*, 155–160. http://dx.doi.org/10.2307/1924423

McDowell, J. M., & Smith, J. K. (1992). The effect of gender-sorting on propensity to coauthor – Implications for academic promotion. *Economic Inquiry*, *30*, 68–82. http://dx.doi.org/10.1111/j.1465-7295.1992.tb01536.x

Meng, Y. (2016). Collaboration patterns and patenting: Exploring gender distinctions. *Research Policy*, *45*, 56–67. http://dx.doi.org/10.1016/j.respol.2015.07.004

Naldi, F., Luzi, D., Valente, A., & Parenti, I. V. (2004). *Scientific and technological performance by gender. Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of ST Systems.* pp. 299–314. ISBN:978-1-4020-2755-0.

Newman, M. (1999). Assortative mixing in networks. *Physical Review Letters*, *89*(20), 208701. http://dx.doi.org/10.1103/PhysRevLett.89.208701

Norris, M., & Oppenheim, C. (2007). Comparing alternatives to the Web of Science for coverage of the social sciences' literature. *Journal of Informetrics*, *1*(2), 161–169. http://dx.doi.org/10.1016/j.joi.2006.12.001

OECD. (2016). *OECD Data Researchers*. https://data.oecd.org/rd/researchers.htm (accessed on July 2016)

Pieters, R., & Baumgartner, H. (2002). Who talks to whom? Intra- and interdisciplinary communication of economics journals. *Journal of Economic Literature*, *40*, 483–509. http://dx.doi.org/10.1257/jel.40.2.483

Raasch, C., Lee, V., Spaeth, S., & Herstatt, C. (2013). The rise and fall of interdisciplinary research: The case of open source innovation. *Research Policy*, *42*(1), 138–1151. http://dx.doi.org/10.1016/j.respol.2013.01.010

Rhoten, D., & Pfirman, S. (2007). Women in interdisciplinary science: Exploring preferences and consequences. *Research Policy*, *36*, 56–75. http://dx.doi.org/10.1016/j.respol.2006.08.001

Rorstad, K., & Aksnes, D. W. (2015). Publication rate expressed by age gender and academic position – A large-scale analysis of Norwegian academic staff. *Journal of Informetrics*, *9*, 317–333. http://dx.doi.org/10.1016/j.joi.2015.02.003

Ruiz-Castillo, J., & Waltman, L. (2015). Field-normalized citation impact indicators using algorithmically constructed classification systems of science. *Journal of Informetrics*, *9*(1), 102–117. http://dx.doi.org/10.1016/j.joi.2014.11.010

Stefaniak, B. (2001). International co-operation in science and in social sciences as reflected in multinational papers indexed in SCI and SSCI. *Scientometrics*, *52*, 193–210. http://dx.doi.org/10.1023/A:1017959605687

Sugimoto, C. R., Lariviere, V., Ni, C., Gingras, Y., & Cronin, B. (2013). Global gender disparities in science. *Nature*, *504*, 211–213. http://dx.doi.org/10.1038/504211a

Testa, J. (2016). *The Thomson Reuters Selection Process*. http://wokinfo.com/essays/journal-selection-process/ (retrieved 18.07.16)

Torres-Salinas, D., & Moed, H. F. (2009). Library Catalog Analysis as a tool in studies of social sciences and humanities: An exploratory study of published book titles in Economics. *Journal of Informetrics*, *3*(1), 9–26. http://dx.doi.org/10.1016/j.joi.2008.10.002

Tsai, C. C., Corley, E. A., & Bozeman, B. (2016). Collaboration experiences across scientific disciplines and cohorts. *Scientometrics*, 1–25. http://dx.doi.org/10.1007/s11192-016-1997-z

Uhly, K. M., Visser, L. M., & Zippel, K. S. (2015). Gendered patterns in international research collaborations in academia. *Studies in Higher Education*, *8*, 1–23. http://dx.doi.org/10.1080/03075079.2015.1072151

Vilela Mendes, R., Araújo, T., & Louçã, F. (2003). Reconstructing an economic space from a market metric. *Physica A*, *323*, 635–650. http://dx.doi.org/10.1016/S0378-4371(03)00014-1

Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics*, *10*(2), 347–364. http://dx.doi.org/10.1016/j.joi.2016.02.003