Scientometrics (2018) 117:953–972
https://doi.org/10.1007/s11192-018-2903-7

Are scientific memes inherited differently from gendered authorship?

Tanya Araújo[1,2] · Elsa Fontainha[1,3]

Received: 9 February 2018 / Published online: 31 August 2018

Abstract
This paper seeks to build upon the previous literature on gender aspects in research collaboration and knowledge diffusion. Our approach adds the meme inheritance notion to traditional citation analysis, as we investigate if scientific memes are inherited differently from gendered authorship. Since authors of scientific papers inherit knowledge from their cited authors, once authorship is gendered we are able to characterize the inheritance process with respect to the frequencies of memes and their propagation scores depending on the gender of the authors. By applying methods that enable the gender disambiguation of authors, missing data on the gender of citing and cited authors is dealt with. Our empirically based approach allows for investigating the combined effect of meme inheritance and gendered transmission. Results show that scientific memes do not spread differently from either male or female cited authors. Likewise, the memes that we analyse were not found to propagate more easily via male or female inheritance.

Keywords Memes · Knowledge diffusion · Gender · Bibliometrics · Research collaboration

JEL Classification O3 · C55 · C89

Introduction

The availability of very large databases has contributed to enlarge the horizons of knowledge in multiple domains, such as social media communication, health, business, finance, economics, invention and innovation, and scientific diffusion and progress. Some recent literature based on big data spans multiple data sources, disciplines and applications, including (1) financial markets forecasting through text-based information from news and

✉ Tanya Araújo
tanya@iseg.ulisboa.pt

[1] ISEG (Lisbon School of Economics and Management), Universidade de Lisboa, Rua do Quelhas, 6, 1200-781 Lisbon, Portugal

[2] REM Research in Economics and Mathematics, UECE Research Unit on Complexity and Economics, Lisbon, Portugal

[3] CEsA GSG Research in Social Sciences and Management, Rua Miguel Lupi, 20, 1249-078 Lisbon, Portugal

social media (Bukovina 2016); (2) investor sentiment research built on machine learning and natural language processing (Curme et al. 2015); (3) analysis of co-occurring terms in opinion dynamics (Banisch et al. 2012; Weichselbraun et al. 2014) and (4) innovation diffusion based on epidemic models (Pulkki-Brännstrom and Stoneman 2013).

Notwithstanding, most analyses of knowledge transmission, channels and mechanisms are based on classical large scientific databases like Google Scholar, PubMed, Scopus and Web of Science, which have been intensely compared, discussed and evaluated (Adriaanse and Rensleigh 2013; Bakkalbasi et al. 2006; Falagas et al. 2008; Harzing and Alakangas 2016; Kulkarni et al. 2009; Meho and Yang 2007).

Besides providing raw data on publications and patents, some large scientific databases publish science metrics, such as Impact Factor per journal, h-index per author and citation scores. Studies using citation indicators have provided insights into scientific performance and impact (Evans 2012), and knowledge dissemination among individuals (Sorenson 2006), firms (Aharonson and Schilling 2016) or regions (Bornmann et al. 2015), as well as, between universities and firms (Azagra-Caro et al. 2016). Citation content and frequency also feed network approaches used in the study of social contagion mechanisms. Research collaboration and co-authorship help to discover patterns of collaboration within scientific communities of authors, inventors or innovators (Araújo and Fontainha 2017).

Kuhn et al. (2014) study the spread of scientific knowledge using Dawkins's concept of meme, the cultural analogy of gene in the context of genetic evolution. Illustrations of a scientific meme as a replicator are provided in The Selfish Gene book by Richard Dawkins: "If a scientist hears, or reads about, a good idea, he passes it on to his colleagues and students. He mentions it in his articles and his lectures. If the idea catches on, it can be said to propagate itself, spreading from brain to brain. If the meme is a scientific idea, its spread will depend on how acceptable it is to the population of individual scientists; a rough measure of its survival value could be obtained by counting the number of times it is referred to in successive years in scientific journals" (Dawkins 1976, p. 194).

Although the idea of memes is not completely original, as Dawkins acknowledges, it has received growing interest ever since. The concept of meme has been explored in several scientific areas. In Economics, Robert Shiller recently drew attention to memes and narratives in his Presidential Address delivered at the American Economic Association meeting: "There is a daunting amount in the scholarly literature about narratives, in a number of academic departments, and associated concepts of memetics, norms, social epidemics, contagion of ideas. While we may never be able to explain why some narratives go viral and significantly influence thinking while other narratives do not [...] We economists should not just throw up our hands and decide to ignore this vast literature" (Shiller 2017, p. 972).

Here, we aim to investigate if scientific memes are inherited differently from gendered authorship. Since authors of scientific papers inherit knowledge from their cited authors, once authorship is gendered (by applying methods that enable the gender disambiguation of citing and cited authors), we are able to characterize the inheritance process with respect to the frequencies of scientific memes and their propagation scores depending on the gender of the authors. Would female inheritance—represented by the citations of female authors—favor the propagation of some specific meme? Likewise, would some particular memes propagate more via male inheritance?

Moreover, our paper seeks to build upon the previous literature about gender aspects in research communication, collaboration and co-authorship (Araújo and Fontainha 2017; Astebro and Thompson 2011; Bozeman and Gaughan 2011; Brooks et al. 2014; Gonzalez-Brambila and Veloso 2007; Sugimoto 2015; Tartari and Salter 2015; Van Rijnsoever and

Hessels 2011; Viana 2013; Ynalvez and Shrum 2011) and scientific outcome impact by gender (Beaudry and Lariviere 2016; Copenheaver 2010; de Melo-Martin 2013; Frietsch et al. 2009; Giuri et al. 2007; Ghiasi 2015; Hunt et al. 2013; Jung and Ejermo 2014; Mauleón and Bordons 2010; Meng 2016; Mihaljevic-Brandt 2016; Okon-Horodynska et al. 2015).

By including gender in the study of knowledge spread and adding a gender perspective to the *'spreading of good ideas, from brain to brain'*, to adopt Dawkins's words, our empirically based research aims to contribute in three ways to the improvement of the understanding of the way knowledge spreads:

- It adds the meme inheritance notion to traditional citation network analysis;
- It accurately identifies the gender of authors, dealing with the issue of missing data in large scientific databases, and
- It sheds some light on the existence of gender-homophily trends in citation choices

There is a broad literature (Granovetter 1973; Sorenson 2006; Krichel and Bakkalbasi 2006; Borgatti et al. 2009; Cainelli et al. 2015) on social relations (social networks included) showing that many social systems create contexts in which homophilic relationships hold. From friendship, co-membership and marriage, several studies have discussed the role the similarity plays in the creation of human relationships. The phenomena of establishing ties with similar individuals have been extensively studied through network approaches, regardless whether similarity is based on age, religion, education, occupation, or gender.

Recent research on the structure of citation networks (Ciotti et al. 2016) presents a method for measuring the similarity between articles through the overlap between the bibliographic lists of references included in these articles (cited papers). One related study is discussed by Ferrer i Cancho et al. (2004) with the definition of a similarity network between articles on linguistic, cognitive and brain networks. There, instead of bibliographies, the similarity between articles is measured on the basis of similar words used in the abstract of the articles. Therefore, the network approach allows for clustering articles on linguistic networks into different modules depending on whether they deal with semantics or functional brain networks.

When the gender aspect is considered, a large scale analysis on gendered authorship (West et al. 2013) based on eight million papers across multiple areas reveals that women are significantly under-represented as authors of single-authored papers. Araújo and Fontainha (2017) arrived to close results when analyzing gender authorship of scientific papers through a network approach. This paper, seeking to build upon the previous literature on gender aspects in research transmission adds to usual citation analysis the meme approach and propagation score methodology. Our computation of the propagation scores of memes characterized by gendered authorships of the citing and cited papers allows for investigating the combined effect of meme inheritance and gendered transmission.

The remainder of this paper is structured as follows: "Gender studies in bibliometrics" section presents a short overview of the literature on bibliometrics. "Data and methodology" section briefly describes some scientific databases calling attention to the lack of information about the gender of citing and cited authors. "Results and discussion" section presents the data and the methodologies used in the paper. "Conclusion" section presents and discusses the results from the empirical analyzes. In the final section conclusions and some promising research avenues are provided.

Gender studies in bibliometrics

In recent years in bibliometric literature, there has been a growing research on gender issues. However, a major limitation exists in large bibliometric databases: the authors' gender is not part of the available information and only indirectly this information can be obtained. When WOS started to make available the first name of the authors (since 2007) it became possible to apply large scale and automatic methods of gender disambiguation and partially overcome this limitation. Different methods have been used: (a) disambiguation based on the first name, available for example in WOS for the recent years (e.g. Lariviere et al. 2013) or Scopus (Elsevier 2017); (b) disambiguation based on the first name together with bibliographic data (Naldi et al. 2004); or (c) disambiguation based on specificities of the family name, as in Polish (Kosmulski 2015) or in Russian names (Paul-Hus et al. 2015).

Disambiguation methods have shown that compared with men, women produce less (Lariviere et al. 2013), do not perform a central role in collaboration networks (Ghiasi 2015), are less frequently the key author (Macaluso et al. 2016), and are cited less frequently. It must be pointed out that these results are strongly affected by the absence of information about the input indicators of research in general (Waltman et al. 2016) and by gender. The productivity or performance of women to be evaluated with rigor, need to compare input indicators (as the number of researchers by gender and time allocated to research) with output measures (as articles published and citations received).

Bibliometric research includes gender issues in two main sets of studies: one has gender themes as the core of the research; the other combines gender variables with other characteristics (e.g. institutions, fields of science) to explain research production, performance and impact.

Research production, performance and impact

There is a gap in the literature using large scale studies about research production, performance and impact by gender. Notwithstanding, there have been some few studies on gender developed at a large scale (Elsevier 2017; Lariviere et al. 2013). Lariviere et al. (2013), study gender disparities in science based on 5,483,841 scientific outputs (research papers and review articles) using web of science (WoS) bibliographic database as source. Their global and cross-disciplinary bibliometric analysis concludes that male authors dominate the global scientific production, and also most of the disciplines including humanities. The report entitled *Gender in the Global Research Landscape* (Elsevier 2017) analysing the publications over two decades (1996–2015), in 27 scientific areas arrives to some similar conclusions. However, the existence of dissimilar and even contradictory outcomes in those two studies (Elsevier 2017; Lariviere et al. 2013) illustrates how can results be affected by data sources, gender disambiguation methods, scientific fields and time frames.

Abramo and co-authors, using Italian large scale databases, have a large and consistent body of research by gender on scientific performance and productivity (Abramo et al. 2009, 2014, 2015), research collaboration (Abramo et al. 2013) and academic career (Abramo et al. 2014). The gender identification is made matching the bibliometric databases with a list of researchers where the field gender is available.

Most of the studies on research production, performance and impact by gender are based on small samples. Nielsen (2017) shows that in the 16 articles on management surveyed, only one (Podsakoff et al. 2008) includes large scale databases. The sample size in the other 15 studies vary between a minimum of 57 articles (Mitchel et al. 2013) and a maximum of 2541 articles (Maliniak et al. 2013). Potthoff and Zimmermann (2017), analysing 917 articles in two communication journals, find gender-homophily trends in citations: male authors cite publications of male authors more frequently than their female colleagues do and the symmetrical also happens. They attribute these results to topical boundaries in science by gender.

Physics has not been focus of many scientometric studies by gender. Paul-Hus et al. (2015) analysing 1,059,939 scientific outputs indexed in Web of Science and published between 1973 and 2012 by authors affiliated to Russian institutions, conclude that in that country, where Mathematics and Physics correspond to a large part of research, female academics are under represented.

Bibliometric databases

Four databases (Web of Science, Scopus, Google Scholar and PubMed), and one repository (arXiv) were searched in order to explore the possibilities of extracting information regarding the gender of both citing and cited authors and the memes found in the abstracts of citing and cited records (papers). The following criteria were adopted to select the datasets to examine in detail: size and accuracy of the data; tracking citation possibilities; down-loading capabilities; and possibility to gather, for each record, at least, its title and abstract.

Several studies have compared the coverage, features, and citation analysis capabilities of GS, PubMed, Scopus and WoS. These comparative studies usually focus on a particular research topic like biomedical information (Falagas et al. 2008), medical journals (Kulkarni et al. 2009), oncology and condensed matter physics (Bakkalbasi et al. 2006) library and information science (Meho and Yang 2007) or environmental sciences (Adriaanse and Rensleigh 2013). Other studies address only the accuracy of one database (Franceschini et al. 2016). This literature, however, fails to systematically review the citation analysis linked with the author full identification.

- Web of Science (WoS): formerly the ISI Web of Knowledge, adopts a selection process for the inclusion of journals in its content coverage (Testa 2016). The most frequent criticisms to WoS are the bias to American-based, English-language journals, failure to completely cover other citation sources (e.g. books) and failure to include citations out of the WoS database. Despite the criticism, WoS is often used worldwide in scientometrics analysis based on information articles and articles citation on a subject (Adriaanse and Rensleigh 2013; Harzing and Alakangas 2016; Sugimoto 2015; Lariviere 2008). From 2007 onwards the reporting of author's name in WoS changed with the inclusion of the full name (given name of the authors). However, the full name of cited authors, i.e., the authors in the bibliographic list of references of each article is not provided.
- Scopus: includes publications from Sciences, Social Sciences and Art and Humanities. Scopus both covers more journals than WoS and provides better coverage of the non-North American sources. The citations of an author and the articles that cite the original article (using Citation Tracker) make it possible to base the analysis of citations on different criteria and enable the researcher to create an exportable spreadsheet of citations. Author names in Scopus can be arranged differently. Consequently, there are an unknown share of authors in the database where the given name is missing.

- Google Scholar (GS): created in 2004, comprises all fields of knowledge and several types of documents. GS neither defines the number of journals covered nor the time span of the database. Thus, for a document with the same title and authorship, it includes all the versions available online.
- Pub Med: an important resource for clinicians and researchers. The dataset includes Medline (1966-present), old Medline (1950–1965), PubMed Central, and other NLM databases. The full name of the authors in the bibliographic list of references of each article is not provided.
- ArXiv: a pre-print archive of working papers in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics. The repository arXiv, where High Energy Physics belongs, was founded in 1991 by Paul Ginsparg, a theoretical physicist at Cornell University, and since then it has received growing interest and use from the scientific community (Ginsparg 2011). A study comparing this pre-publication repository with publication databases has been carried out by Bar-Ilan (2014). There are several studies about the research impact of the material in the repository of arXiv namely using citation analysis (Brody 2006; Davis and Fromerth 2007; Evans 2012; Goldberg 2015; Haque and Ginsparg 2010). Other studies use arXiv to identify trends and build the agenda for future research in multiple scientific domains (Haque and Ginsparg 2010; Lariviere 2008).

From name to gender

In fact, the two major bibliographic databases, web of science (WoS) and Scopus, both of which cover several scientific domains and many types of scientific outputs, do not include the information needed to answer our research questions. The large databases and repositories with scientific outputs, as well as, most of the repositories, independently of the coverage (by domain, period or type of document) and citation search strategies, produce quantitatively and qualitatively different citation material.

The situation is worse with regard to the transmission of knowledge (from citing to cited author), because the databases provide neither information on citations by gender nor the full name of the cited author, i.e., authors in the bibliographic list of references. Thus, given this lack of information, the only way to overcome these weaknesses is to obtain the information from the first name or the family name of the contributors concerned.

In brief, some databases include the full name of the authors, which enables, at least partially, the identification of his/her gender. However, the bibliographic list of references (citations) in each article does not provide the full name of the cited authors. This lack of information is solved in the present research by using the dataset provided by stanford network analysis platform (SNAP) together with the repository file `nam-dict.txt` (Michael 2008) for the gender disambiguation of authors.

Stanford network analysis platform (SNAP) is a general purpose network analysis and graph mining library. Among the Stanford Large Network Dataset Collection, we were able to download the dataset recorded from the repository arXiv: the `hep-th High Energy Physics` (Leskovec and Sosič 2014). The detailed description of this dataset is presented in the next section.

Data and methodology

We used the dataset recorded from the repository arXiv, the `hep-th High Energy Physics - Theory` and provided by Jure Leskovec at Stanford Large Network Dataset Collection (Leskovec and Sosič 2014). The data covers papers in the period from January 1993 to April 2003 (124 months) within a dataset of 29,555 papers and 352,807 links.

The available dataset is organized in three files, two of which were the main source of the work herein presented:

(1) `cit-HepTh-abstracts`	includes:		
	• `Paper` *id*	• `From`	• `Authors`
	• `Title`	• `Date`	• `Abstract`
(2) `cit-HepTh` provides:	• List of	edges	from `Paper` *i* to `Paper` *j*

Each directed link in the citation network `cit-HepTh` indicates that `paper` *i* cites `paper` *j*. If a paper cites, or is cited by, a paper outside the dataset, the list does not contain any information about this.

For each paper in the dataset, the field `Abstract` includes the corresponding abstract of the paper. The field `From`, a metadata field for arXiv submissions is automatically filled with the information on who submits the article. The field `Authors` provides the full name of the authors in 66.6% of the observations. Since most of the papers in this dataset comprise the first (given) name of the authors, we were able to classify the citing and cited authors by gender.

Gender disambiguation of authors

Bibliometric databases have been the main sources for the study of knowledge spread. However, due to the lack of information on the gender of the authors, the study of the scientific production by gender has been frequently limited to surveys or case studies (Frietsch et al. 2009). When the number of observations is low, gender allocation is done manually. Automatically allocating gender to researchers depends on the availability of: (1) databases with gender information that allow matching by researcher name or code. For example, Abramo et al. (2013) combine data from WoS with data from the Italian Ministry of Education, Universities and Research; (2) databases with a list of male and female given names in different languages, as for instance, the database built under a EU project *Improving Human Research Potential and the Socio-Economic Knowledge Base* (Naldi and Parenti 2002; Naldi et al. 2004) and the method used in a recent report from Elsevier (2017); (3) language specific characteristics that allow for systematically allocating gender from each researcher's family name as, for example, Polish names allow to allocate gender from the family name (Kosmulski 2015).[1]

This research adopts the methodology mentioned in (2) and uses a GitHub file (Michael 2008) as the data source for the gender disambiguation of authors. The steps are the following:

[1] Most of the Asian researchers when publishing in English-language journals have to adopt a phonetic version of the given and family names and this creates ambiguities to the authors' gender attribution (Qiu 2008).

Table 1 Paper-centered information from the original dataset after gender disambiguation of authors

	All	With gender
Number of papers	29,555	19,696
Number of citations	339,741	197,378
Number of citing papers (Ci)	25,058	17,230
Number of cited papers (Ce)	23,180	15,596
Size of ($Ci \bigcup Ce$)	27,770	19,153
Size of ($Ci \bigcap Ce$)	20,468	13,673

- From the GitHub repository a file (`nam-dict.txt`) comprising 48.258 (first) names is downloaded.
- From each record (paper) in the `cit-HepTh-abstracts`, the first name recorded in the field `Authors` is extracted.
- When a first name is missing or contains just initials, the first name in the field `From` is extracted if, in the same record, there is a family name in `Authors` that is also in `From`. It happened in 8102 records (2.7%) being an alternative source to get the first name of an author.
- Each extracted first name is searched in the file `nam-dict.txt`.
- Each search returns either: male (M), female (F), both (=), undefined (?) or empty (`e`).

In so doing, we are able to extract one first name in 24,171 to which gender was confidently assigned in 19,696 (81.4%) papers. The remaining papers 4,475 (18.6%) provide names not in the GitHub package (marked with the signs "?" and "e"), or names associated with both genders (marked with =). Tables 1 and 2 in "Sample characteristics" section summarize these results.

Author selection

The assignment of author ordering differs through and inside scientific disciplines (Abramo and D'Angelo 2017). In biomedical sciences, for instance, the main author use to be the last listed. In high energy physics, authors go by alphabetical order (Waltman 2012).

Figure 1 shows the distribution of the number of authors per publication. More than 20,000 papers have at most two authors and more than 10,000 are single-authored papers. Besides simplification, focusing just on the first author is partially due to the use of the first name in the field `From` when the first name is missing (or abbreviated) in the field `Authors`. The field `From`, a metadata field for arXiv submissions is automatically filled with the information on who submits the article.

The selection of one single author or a subgroup of authors among the total number of authors of a given article is done in other bibliometric studies. For example, Potthoff and Zimmermann (2017) analyze only the first-named authors and build the variable gender composition for each article based on the gender of the first- and second-named author (p. 1054 and 1057). In our research the first-named author (preferred designation compared to 'first author') is selected for the analysis. He or she is not necessarily the representative author or the most relevant author because in High Energy Physics there is a predominance of applying alphabetic order (Waltman 2012). The use of alphabetic order among the authors in the non-single authored papers was tested and proved: in our sample, authors go by alphabetic order in 90.4% of the papers with at least two authors. Consequently, it can be assume that when only the first-named author is selected, the selection tends to be

Table 2 Author-centered information after the gender disambiguation of authors

	All	Female	Male	Missing
Number of 1st authors	14,099	1079	8751	4269
Number of 2nd authors	6496	687	4200	1609
% of 1st authors citing by gender	100	11.1	88.8	–
% of 1st authors cited by gender	100	10.9	89.1.0	–
% of 2nd author by gender		10.4	89.6	–

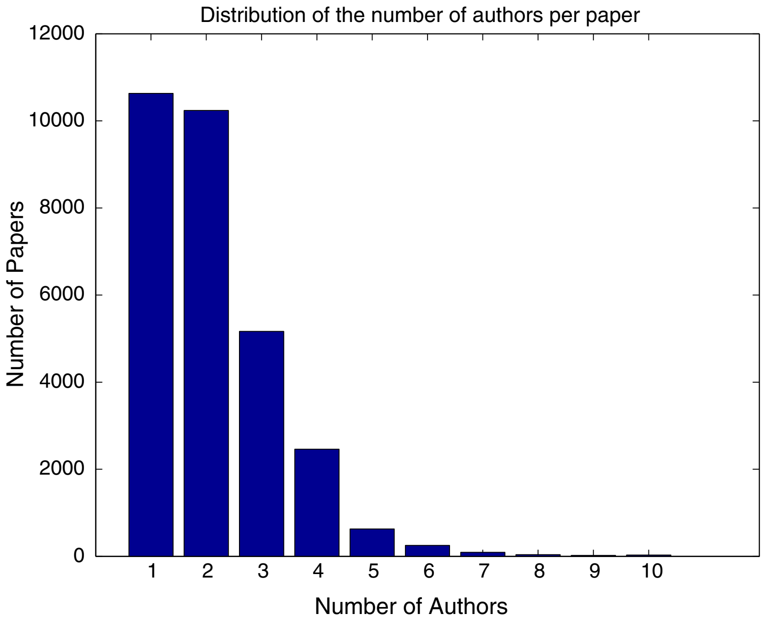

Fig. 1 The distribution of the number of authors per paper

random concerning the role performed by each author in a non-single author article and the results are independent from the role performed by a specific co-author.[2]

Sample characteristics

Tables 1 and 2 display an overview of the basic information compiled from arXiv: the `hep-th High Energy Physics` (Leskovec and Sosič 2014) dataset after the gender disambiguation of authors. Author gender was accurately assessed for 81.4% of the papers. As earlier mentioned, the loss of information in the process of disambiguation is in line with other studies (Beaudry and Lariviere 2016; Lariviere 2008). A *gendered paper* is a paper that enables the gender disambiguation of at least its first author. A *gendered author*

[2] We repeated the calculations considering just the papers with a single author (both for citing and cited papers). The results are quite the same and they are available for interested readers upon request.

is an author for whom the database enables the identification of his/her gender. A *gendered link* is a link between two gendered papers. 58% of the citations are gendered links. There is an overlap of approximately 2/3 of the papers that are both citing and cited papers in the citation network. This ratio also applies when we consider just the papers with gendered authors, as the values in Table 1 show.

Table 2 provides author-centered information considering just the first author of each paper, which corresponds to 9830 unique authors. Its distribution by gender yields 1079 female and 8751 male authors. Self-citations are not considered. The percentage of female authors in the citing papers is 11.1 and in the cited papers is 10.9

The difference between the percentage of female cited and male cited authors seems to mirror the universe of the publications' authorship (citing authors) by gender. Because the cited papers correspond by definition to a period of time before the citing papers, and the proportion of female researchers for the subject area Physics and Astronomy has tended to increase in developed countries (Elsevier 2017), the slight difference found merely reflects a potential number of citable papers authored by women that is smaller than that of the citing papers.

Even though the main topic of this research is not centered on Citation Networks, from the whole set of unique authors (14,099) and the set of 197,378 citations in our dataset, we are able to define a citation network of authors, comprising 9830 gendered authors and 197,378 directed links between each pair of citing-cited (gendered) authors. As in citation networks of authors, a pair of nodes i, j is connected if and only if the author i cites the author j.

Figure 2 shows a connected sub-network of the whole citations network. This network comprises 858 authors (citing and cited) and just 3130 links. In both graphs of Fig. 2, the networks comprise the same nodes and links but nodes are sized and colored differently. In the network in the left the nodes are sized and colored accordingly to the number of their out-going links while in the network in the right, the size and the darkness of the nodes are proportional to the number of their out-degrees. Looking at these graphs, one sees that the distribution of the in-coming links per author is much more unequal than the distribution of the out-going ones (19.98). The distribution of the citations made is much more

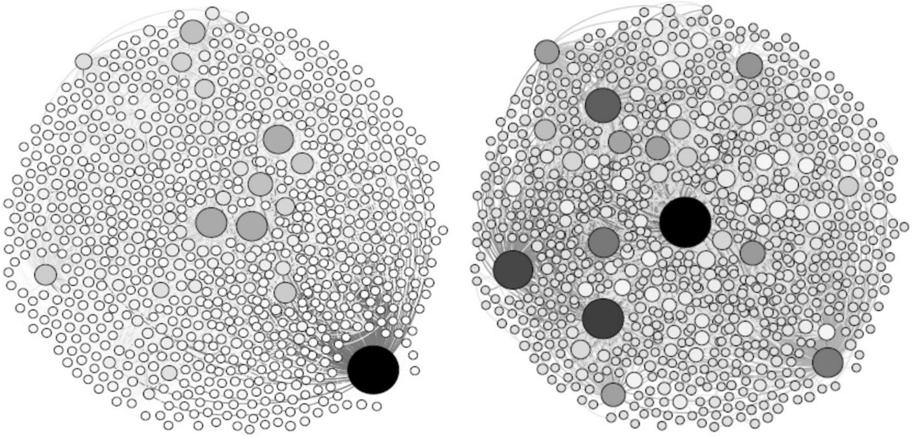

Fig. 2 A connected sub-network (858 nodes, 3130 links) of the whole network of citations. In the left, nodes are sized and colored proportionally to their in-degrees. In the right, the size and the darkness of the nodes are proportional to their out-degrees

homogeneous than the distribution of the citations received, although on average, the number of in-coming links (20.08) is close to the number of out-going ones, the latter are much less equally distributed. The two graphs in Fig. 2 show the remarkable difference between the distribution of the in-degrees and out-degrees. In the whole network, there is an author that receives more than 3000 citations. A much more balanced distribution characterizes the citations made, where the most citing author does not go beyond 467 citations.

Memes selection

Following the approach of Kuhn et al. (2014), our research is driven by the characterization of the propagation (or inheritance) mechanism of memes and not just by their frequency of occurrence as is usual in citation analysis.

A first step into this direction is the selection of a sub set of memes among the whole set of the most frequently occurring words in the entire set of 29,555 papers. Our meme selection process starts by using the word-counting procedure of Voyant Tools (Sinclair and Rockwell 2016). Voyant Tools allows for defining a list of words to be excluded from the word-counting procedure (*stopwords*). Typically, a *stopword* list contains functional words that do not carry much meaning, such as determiners and prepositions ("in", "to", "from", among others). Table 3 shows 40 memes selected among the most frequently occurring words in the abstracts (without *stopwords*) and ranked by frequency of occurrence. It also shows the frequency of each selected meme computed from both the 29,555 papers and from the 19,696 gendered papers.

Table 3 The absolute frequency of 40 selected memes from the frequently occurring words in the abstracts of the 29,555 papers

Rank	Meme	F_m	Rank	Meme	F_m
1	Space	9249	2	Gauge	8082
3	String	7517	4	Quantum	6275
5	Symmetry	5682	6	Brane	5153
7	Mass	5082	8	Gravity	4621
9	Group	4600	10	Conformal	3389
11	Potential	3331	12	Spin	2604
13	Hole	2395	14	Supersymmetry	2220
15	Supergravity	2118	16	Topological	2079
17	Phase	2068	18	Abelian	2034
19	Magnetic	1983	20	Manifold	1967
21	Matter	1829	22	Spacetime	1812
23	Vacuum	1802	24	Coupled	1795
25	Tensor	1763	26	Massless	1654
27	Renormalization	1418	28	Cosmological	1393
29	Gravitational	1362	30	Bosonic	1352
31	Chern	1277	32	Temperature	1172
33	Lattice	1033	34	Discrete	1023
35	Fermionic	981	36	Relativistic	932
37	Superconformal	752	38	Singularity	727
39	Cohomology	465	40	Hierarchy	464

The memes selected correspond to the most frequent words in our sample that carry a specific meaning in the field of High Energy Physics. Many frequent words like “theory”, “dimensional” or “field” were excluded since they are not enough specific, occurring also frequently in papers found in other scientific areas. These words were excluded together with functional words because we are interested in the thematic similarities of the papers as opposed to, for instance, stylistic similarities between different authors. Therefore, we disregarded functional words, as well as, those words common to many other scientific areas.

Figure 3 shows the relative frequencies (the ratio of papers carrying the meme in each subset of papers) of each selected meme in Table 3. The relative frequencies are computed from both the 29,555 papers and from the subset of the 19,696 gendered papers. There are some small differences in the values of the relative frequencies depending on whether they are computed from either the 29,555 papers (f_m) or from the subset of the 19,696 gendered papers (f_m^g). Those differences are, on average, smaller than 5% of f_m and in two thirds of the memes the value of f_m^g is greater than the corresponding f_m value, meaning that, the relative frequencies of 2 / 3 of the selected memes slightly increase when computed from the gendered papers. The vertical dashed line in Fig. 3 points out the 15 memes whose frequency of occurrence in the subset of gendered papers is above 8%. In the next section,

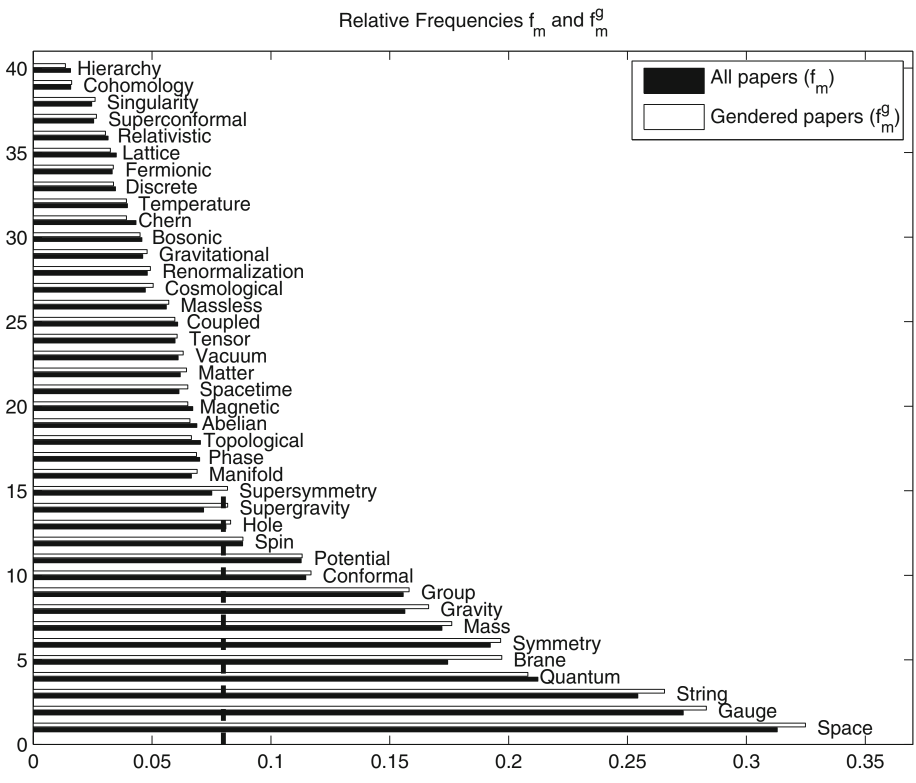

Fig. 3 The relative frequency of occurrence of 40 memes computed from all 29,555 papers (f_m) and from 19,696 gendered papers (f_m^g)

we compute the propagation score of these 15 memes and discuss its relation with the frequency of occurrence.

Results and discussion

Since authors of scientific papers inherit knowledge from their cited authors and once authorship is gendered, our research questions can be rephrased:

- Is the frequency and propagation of a meme (from paper cited to paper citing) influenced by the gendered cited paper?
- Do the selected memes spread differently from either male or female cited authors?

To answer these questions we characterize the inheritance process with respect to the frequencies of memes and their propagation scores depending on the gendered authorship of the cited papers.

Departing from such a gender-oriented perspective and restricting our sample to the set of 19,696 gendered papers, two indicators are computed for each selected meme: the relative frequency and the propagation score (Kuhn et al. 2014).

As already mentioned, the relative frequency of a meme computed from the set of (19,696) gendered papers (f_m^g) is the ratio of papers carrying the meme in this subset. The propagation score P_m^g is given by:

$$P_m^g = \frac{d_{m \to m} / d_{\to m}}{d_{m \nrightarrow m} / d_{\nrightarrow m}} \tag{1}$$

where $d_{m \to m}$ is the number of papers that carry the meme m and cite at least one paper carrying this meme, while d_m is the number of all papers (meme carrying or not) that cite at least one paper that carries the meme m. Following Kuhn et al. (2014), we also compute $d_{m \nrightarrow m}$ as the number of papers that carry the meme m and do not cite a paper carrying this meme, and $d_{\nrightarrow}m$ is the number of all papers (meme carrying or not) that do not cite a paper that carries the meme m.

Since in P_m^g, g stands for gendered, its computation is made from the citation network of (206,405) gendered links. When computing the propagation score for each specific gender (P_m^F and P_m^M), we constrain the subsets of links being considered so that the cited papers conform to each specific gender. Therefore, in computing the female (male) propagation score of a meme P_m^F (P_m^M), the terms $d_{m \to m}$, $d_{\to m}$,$d_{m \nrightarrow m}$ and $d_{\nrightarrow m}$ account just for the cited papers of a female (male) author.

Figures 4 and 5 show, respectively, the values of the relative frequencies (f_m^F and f_m^M) and propagation scores (P_m^F and P_m^M) of the 15 memes whose frequency of occurrence in the subset of gendered papers is above 8%. The values in front of the bars correspond to the normalized difference (δ) between P_m^M and P_m^F, being $\delta = \frac{P_m^M - P_m^F}{P_m^F + P_m^M}$. Therefore, negative values indicate that $P_m^M < P_m^F$ and the positive values correspond to the opposite situation.

The results confirm the almost absence of any difference between female and male transmission of memes. The only noteworthy difference in the propagation score by gender concerns the value obtained for the meme 'Spin'. In this specific case, the propagation score via male inheritance is weaker than via female inheritance, being $\delta = -12$. In the other 14 cases, results confirm the almost absence of any difference between female and male transmission of memes.

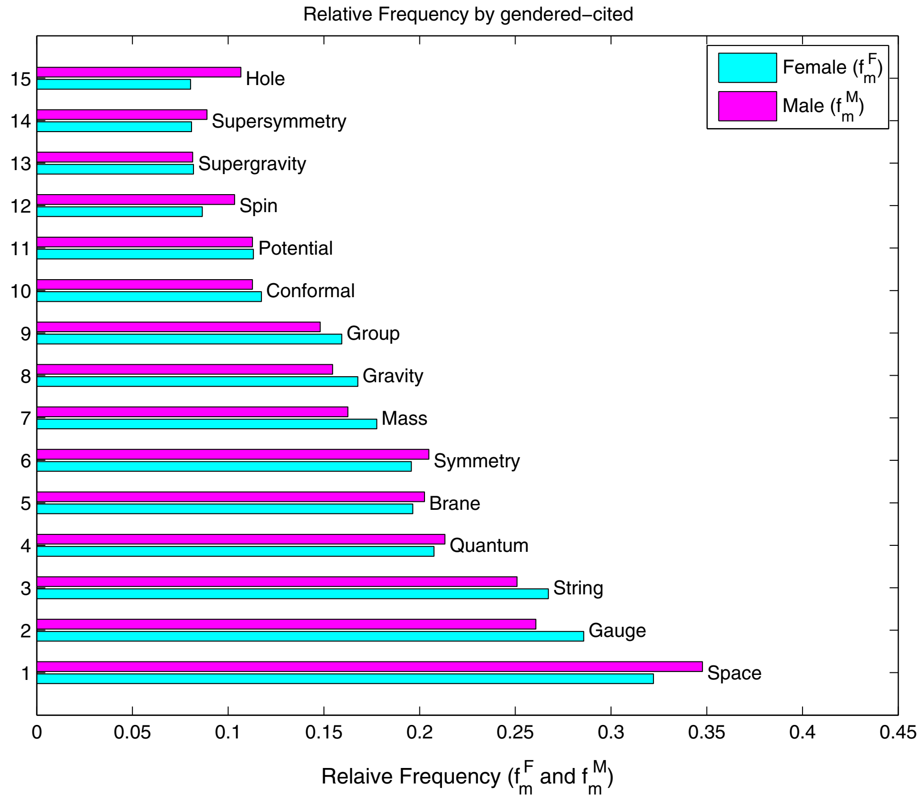

Fig. 4 The relative frequency of memes (f_m^F and f_m^M) in gendered papers

Figure 6 shows a scatter plot where the coordinates of each 15 meme is given by its relative frequency (f_m^g) and propagation score (P_m^g). When the relative frequency and propagation score are plotted against each other, our results are in line with the outcomes presented by Kuhn et al. (2014), showing that less frequent memes tend to propagate more (via citation links). A possible reason for such a simple relation between the relative frequencies and propagation scores of scientific memes may rely on the fact that the less frequent ones are presumably more informative and therefore occur less often. Likewise, functional words—such as determiners and prepositions—carrying less meaning, occur very frequently and therefore occupy the most central positions in linguistic (co-occurrence) networks (Zamora-López et al. 2011; Araújo and Banisch 2016).

Computing the correlation coefficient between the values of f_m^g and P_m^g for the set of gendered papers yields -0.58. As the propagation score of a meme captures how interesting it is for the scientific community, our results confirm that being interesting is inversely related to occurring frequently. The scatter plot in Fig. 6 shows that such a simple relation holds when citation ties are gendered.

Not surprisingly and given that information on the gender of the authors that one cites is usually missing, the transmission of memes are free of gender-homophily trends in citation choices.

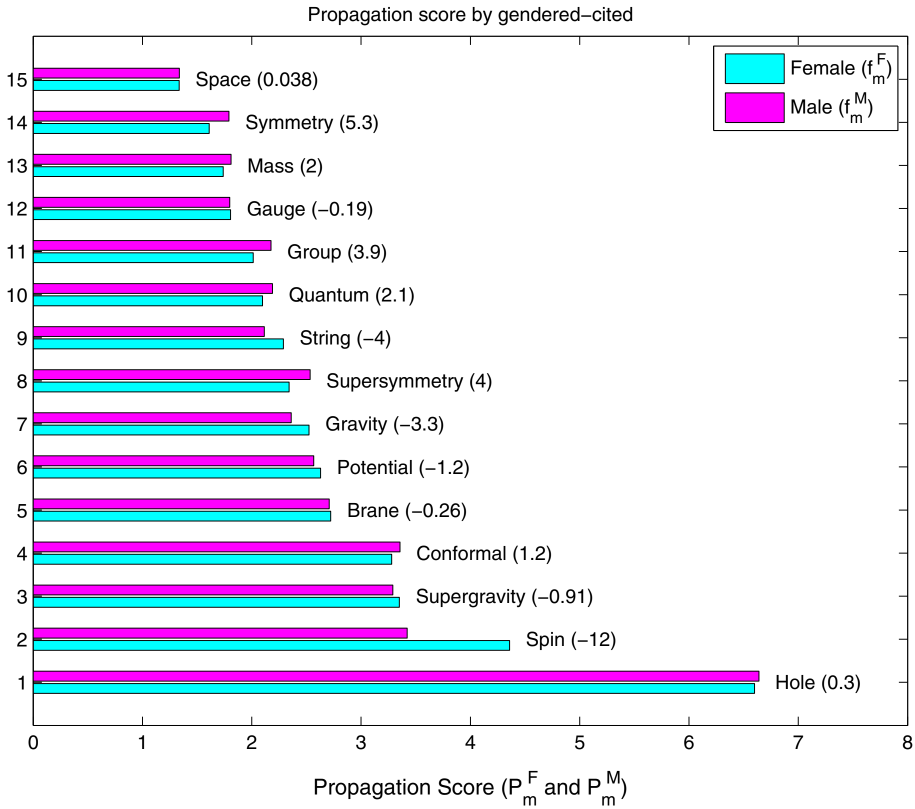

Fig. 5 The propagation score of memes (f_m^F and f_m^M) by gendered cited

Here, we add to usual citation analysis the memes approach and propagation score methodology. Our computation of the propagation scores of memes characterized by the gendered authorship of the citing and cited papers allows for investigating the combined effect of meme inheritance and gendered transmission. In so doing, our results show that the propagation of the selected memes does not seem to be influenced by the gendered authorship. The selected memes do not spread differently from either male or female cited authors. Neither female or male inheritance seems to favor the propagation of any of the selected memes. Likewise, with a single exception, the memes that we analyzed were not found to propagate more easily via male or female inheritance.

Conclusion

Our approach adds the meme inheritance notion to traditional citation analysis, as we investigate if scientific memes are inherited differently from gendered authorship. Results reveal that the inheritance process does not differ by gender. The descriptive analysis suggests the absence of any gender-homophily trend in citation ties. The empirical analysis also shows that there is a very unbalanced scientific output by gender in the scientific domain under analysis. Women represent about $\frac{1}{10}$ of the authorship outputs. Moreover, our results are in line with the results presented in reference (Kuhn et al. 2014), confirming that

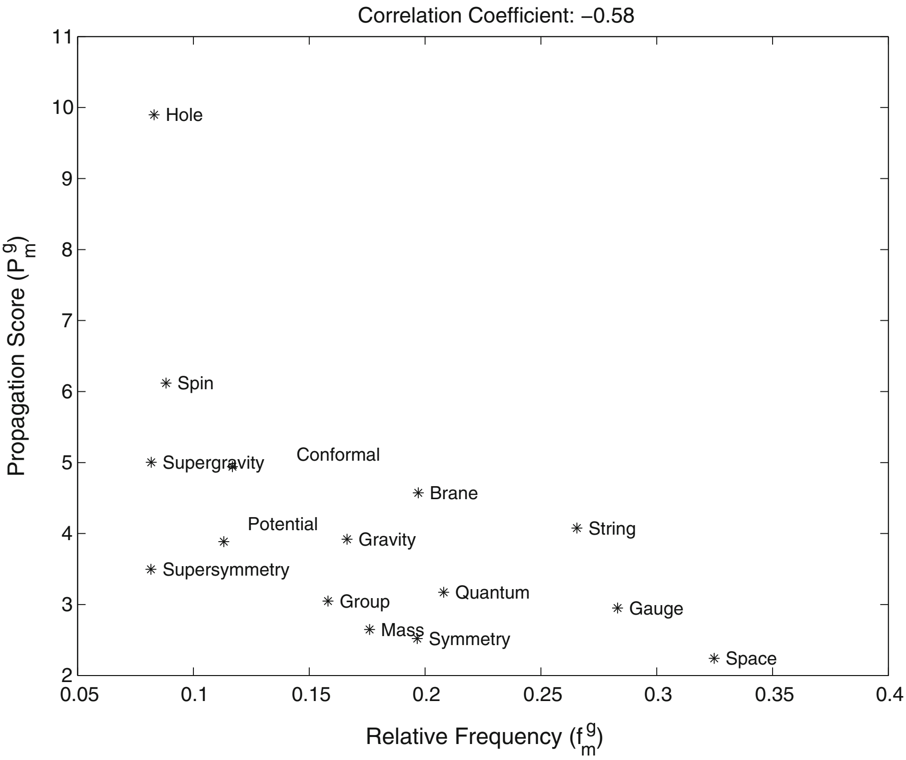

Fig. 6 The relative frequency (f_m^g) and propagation score (P_m^g) of gendered papers plotted against each other

there is a simple relation between the frequency of occurrence of a scientific meme and its propagation score via citation links. Here we show that such a simple relation also holds when citation ties are gendered.

Concerning citation analysis and sciencitometrics, this paper goes a step further investigating the interplay of memes transmission and gendered authorship. The methodology can be useful for academics conducting citation studies and knowledge diffusion analyses. For big data developers, owners, editors, administrators, and funding agencies, the present study also enlarges the horizons of knowledge production and dissemination. In particular, not only are the owners of big databases in a strategic position, but they also have the resources to develop new tools to deal with the lack of information on gender. In the future, when the big bibliometric databases start to include it as a regular procedure, this study can be replicated on a broader scope, free of missing data.

Future research work is planned to further approach citation networks of gendered authors. Following the work of Ciotti et al. (2016) we envision the application of our gender-oriented perspective to the definition of networks of authors based on the overlap between their common references. Therefore, the network approach might allow for clustering gendered authors into different groups depending on multiple characteristics of their bibliographic references. Moreover, applying well-known statistical tools inspired by network studies in other domains, may bring important contributions to the study of networks of scientific collaboration. We envision that, the finding of structural differences between citation networks of different types may be indicative of their usefulness in a more applied context as tools for knowledge diffusion and transfer.

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 R. Vilela Mendes for providing help in the identification of important physics concepts. The research reported in this paper contributes methodologically to one of the case studies (Mapping Gender in Research: the case of Portugal) of the PLOTINA project (Promoting gender balance and inclusion in research, innovation and training). PLOTINA project received funding from the European Union's Horizon 2020 research and innovation programme, under Grant Agreement N. 666008 (www.plotina.eu). The views and opinions expressed in this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.

References

Abramo, G., Cicero, T., & D'Angelo, C. A. (2015). Should the research performance of scientists be distinguished by gender? *Journal of Informetrics*, *9*(1), 25–38. https://doi.org/10.1016/j.joi.2014.11.002.

Abramo, G., & D'Angelo, C. A. (2017). Does your surname affect the citability of your publications? *Journal of Informetrics*, *11*(1), 121–127. https://doi.org/10.1016/j.joi.2016.12.003.

Abramo, G., D'Angelo, C. A., & Caprasecca, A. (2009). Gender differences in research productivity: A bibliometric analysis of the Italian academic system. *Scientometrics*, *79*(3), 517–539. https://doi.org/10.1007/s11192-007-2046-8.

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

Abramo, G., D'Angelo, C. A., & Rosati, F. (2014). Career advancement and scientific performance in universities. *Scientometrics*, *98*(2), 891–907. https://doi.org/10.1007/s11192-013-1075-8.

Adriaanse, L. S., & Rensleigh, C. (2013). Web of Science, Scopus and Google Scholar: A content comprehensiveness comparison. *The Electronic Library*, *31*(6), 727–744. https://doi.org/10.1108/EL-12-2011-0174.

Aharonson, B. S., & Schilling, M. A. (2016). Mapping the technological landscape: Measuring technology distance, technological footprints, and technology evolution. *Research Policy*, *45*(1), 81–96. https://doi.org/10.1016/j.respol.2015.08.001.

Araújo, T., & Banisch, S. (2016). *Multidimensional analysis of linguistic networks, towards a theoretical framework for analyzing complex linguistic networks* (pp. 107–131). Berlin: Springer. https://link.springer.com/chapter/10.1007.

Araújo, T., & Fontainha, E. (2017). The specific shapes of gender imbalance in scientific authorships: A network approach. *Journal of Informetrics*, *11*(1), 88–102. https://doi.org/10.1016/j.joi.2016.11.002.

Astebro, T., & Thompson, P. (2011). Entrepreneurs, jacks of all trades or hobosfi. *Research Policy*, *40*(5), 637–649. https://doi.org/10.1016/j.respol.2011.01.010.

Azagra-Caro, J. M., Barberá-Tomás, D., & Edwards-Schachter & Tur, E. M. (2016). Dynamic interactions between university-industry knowledge transfer channels: A case study of the most highly cited academic patent. *Research Policy*. https://doi.org/10.1016/j.respol.2016.11.011.

Bakkalbasi, N., Bauer, K., Glover, J., & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries*, *3*(1), 7. https://doi.org/10.1186/1742-5581-3-7.

Banisch, S., Lima, R., & Araújo, T. (2012). Agent based models and opinion dynamics as Markov chains. *Social Networks*, *34*(4), 549–561. https://doi.org/10.1016/j.socnet.2012.06.001.

Bar-Ilan, J. (2014). Astrophysics Publications on arXiv, Scopus and Mendeley: A case study. *Scientometrics*, *100*(1), 217–225. https://doi.org/10.1007/s11192-013-1215-1.

Beaudry, C., & Lariviere, V. (2016). Which gender gap? Factors affecting researchers' scientific impact in science and medicine. *Research Policy*, *45*(9), 1790–1817. https://doi.org/10.1016/j.respol.2016.05.009.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, *323*(5916), 892–895. https://doi.org/10.1126/science.1165821.

Bornmann, L., Wagner, C., & Leydesdorff, L. (2015). BRICS countries and scientific excellence: A bibliometric analysis of most frequently cited papers. *Journal of the Association for Information Science and Technology*, *66*(7), 1507–1513. https://doi.org/10.1002/asi.23333.

Bozeman, B., & Gaughan, M. (2011). How do men and women differ in research collaborationsfi an analysis of the collaborative motives and strategies of academic researchers. *Research Policy*, *40*(10), 1393–1402. https://doi.org/10.1016/j.respol.2011.07.002.

Brody, T. (2006). Earlier web usage statistics as predictors of later citation impact. *Journal of the American Society for Information Science and Technology*, *57*(8), 1060–1072. https://doi.org/10.1002/asi.20373.

Brooks, C., Fenton, E. M., & Walker, J. T. (2014). Gender and the evaluation of research. *Research Policy*, *43*(6), 990–1001. https://doi.org/10.1016/j.respol.2013.12.005.

Bukovina, J. (2016). Social media big data and capital markets—An overview. *Journal of Behavioral and Experimental Finance*, *11*, 18–26. https://doi.org/10.1016/j.jbef.2016.06.002.

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. https://doi.org/10.1007/s11192-014-1421-5.

Ciotti, V., Bonaventura, M., Nicosia, V., Panzarasa, P., & Latora, V. (2016). Homophily and missing links in citation networks. *EPJ Data Science*, *5*(1), 7. https://doi.org/10.1140/epjds/s13688-016-0068-2.

Copenheaver, C. A. (2010). Lack of gender bias in citation rates of publications by dendrochronologists: What is unique about this discipline? *Tree-Ring Research*, *66*(2), 127–133. https://doi.org/10.3959/2009-10.1.

Curme, C., Stanley, H. E., & Vodenska, I. (2015). Coupled network approach to predictability of financial market returns and news sentiments. *International Journal of Theoretical and Applied Finance*, *18*(7), 1–26. https://doi.org/10.1142/S0219024915500430.

Davis, P. M., & Fromerth, M. J. (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? *Scientometrics*, *71*(2), 203–215. https://doi.org/10.1007/s11192-007-1661-8.

Dawkins, R. (1976). *The selfish gene*. Oxford Landmark Science. ISBN-13: 978-0198788607

de Melo-Martin, I. (2013). Patenting and the gender gap: Should women be encouraged to patent more? *Science and Engineering Ethics*, *19*(2), 491–504. https://doi.org/10.1007/s11948-011-9344-5.

Elsevier (2017). *Gender in the global research landscape—analysis of research performance through a gender lens across 20 years, 12 geographies, and 27 subject areas*. Elsevier. https://www.elsevier.com/research-intelligence. Accessed 1 Dec 2017.

Evans, T. S. (2012). Universality of performance indicators based on citation and reference counts. *Scientometrics*, *93*(2), 473–495. https://doi.org/10.1007/s11192-012-0694-9.

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google scholar: Strengths and weaknesses. *The FASEB journal*, *22*(2), 338–342. https://doi.org/10.1096/fj.07-9492LSF.

Ferrer i Cancho, R., Solé, R., & Kohler, R. (2004). Patterns in syntatic dependency networks. *Physical Review E*, *69*, 051915. https://doi.org/10.1103/PhysRevE.69.051915.

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2016). The museum of errors/horrors in scopus. *Journal of Informetrics*, *10*(1), 174–182. https://doi.org/10.1016/j.joi.2015.11.006.

Frietsch, R., Haller, I., Funken-Vrohlings, M., & Grupp, H. (2009). Gender-specific patterns in patenting and publishing. *Research Policy*, *38*(4), 590–599. https://doi.org/10.1016/j.respol.2009.01.019.

Ghiasi, G. (2015). On the compliance of women engineers with a gendered scientific system. *PLoS ONE*, *10*(12), 19. https://doi.org/10.1371/journal.pone.0145931.

Ginsparg, P. (2011). arXiv at 20. *Nature*, *476*(7359), 145–147. https://doi.org/10.1038/476145a.

Giuri, P., Mariani, M., Brusoni, S., Crespi, G., Francoz, D., Gambardella, A., et al. (2007). Inventors and invention processes in Europe: Results from the PatVal-EU survey. *Research policy*, *36*(8), 1107–1127. https://doi.org/10.1016/j.respol.2007.07.008.

Goldberg, S. R. (2015). Modelling citation networks. *Scientometrics*, *105*(3), 1577–1604. https://doi.org/10.1007/s11192-015-1737-9.

Gonzalez-Brambila, C., & Veloso, F. M. (2007). The determinants of research output and impact: A study of Mexican researchers. *Research Policy*, *36*(7), 1035–1051. https://doi.org/10.1016/j.respol.2007.03.005.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380. https://sociology.stanford.edu/sites/default/files/publications/the_strength_of_weak_ties_and_exch_wgans.pdf.

Haque, A. U., & Ginsparg, P. (2010). Last but not least: Additional positional effects on citation and readership in arXiv. *Journal of the American Society for Information Science and Technology*, *61*(12), 2381–2388. https://doi.org/10.1002/asi.21428.

Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, *106*(2), 787–804. https://doi.org/10.1007/s11192-015-1798-9.
Hunt, J., Jean-Philippe, G., Herman, H., & Munroe, D. (2013). Why are women underrepresented amongst patentees? *Research Policy*, *42*(4), 831–843. https://doi.org/10.1016/j.respol.2012.11.004.
Jung, T., & Ejermo, O. (2014). Demographic patterns and trends in patenting: Gender, age, and education of inventors. *Technological Forecasting and Social Change*, *86*, 110–124. https://doi.org/10.1016/j.techfore.2013.08.023.
Kosmulski, M. (2015). Gender disparity in polish science by year (1975–2014) and by discipline. *Journal of Informetrics*, *9*(3), 658–666. https://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. http://hdl.handle.net/10760/7406.
Kuhn, T., Matjaz, P., & Dirk, H. (2014). Inheritance patterns in citation networks reveal scientific memes. *Physical Review X*, *4*(4), 041036. https://doi.org/10.1103/PhysRevX.4.041036.
Kulkarni, A. V., Aziz, B., Shams, I., & Busse, J. W. (2009). Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. *Jama*, *302*(10), 1092–1096. https://doi.org/10.1001/jama.2009.1307.
Lariviere, V. (2008). Long-term variations in the aging of scientific literature: From exponential growth to steady-state science (1900–2004). *Journal of the American Society for Information Science and Technology*, *59*(2), 288–296. https://doi.org/10.1002/asi.20744.
Lariviere, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Global gender disparities in science. *Nature*, *504*(7479), 211–213. https://doi.org/10.1038/504211a.
Leskovec, J. & Sosič, R. (2014). SNAP: A general-purpose network analysis and graph-mining library. ACM Transactions on Intelligent Systems and Technology, *8*. arXiv:1606.07550.
Macaluso, B., Larivire, V., Sugimoto, T., & Sugimoto, C. R. (2016). Is Science built on the shoulders of women? A study of gender differences in contributorship. *Academic Medicine*, *91*(8), 1136–1142. https://doi.org/10.1097/ACM.0000000000001261.
Maliniak, D., Powers, R., & Walter, B. F. (2013). The gender citation gap in international relations. *International Organization*, *67*(04), 889–922. https://doi.org/10.1017/S0020818313000209.
Mauleón, E., & Bordons, M. (2010). Male and female involvement in patenting activity in Spain. *Scientometrics*, *83*(3), 605–621. https://doi.org/10.1007/s11192-009-0131.
Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, *58*(13), 2105–2125. https://doi.org/10.1002/asi.20677.
Meng, Y. (2016). Collaboration patterns and patenting: Exploring gender distinctions. *Research Policy*, *45*(1), 56–67. https://doi.org/10.1016/j.respol.2015.07.004.
Michael, J. (2008). SCCS: nam-dict.txt. https://raw.githubusercontent.com/cstuder/genderReader/master/gender.c/nam-dict.txt. Accessed 12 Dec 2016.
Mihaljevic-Brandt, H. (2016). The effect of gender in the publication patterns in mathematics. *PLoS ONE*, *11*(10), 23. https://doi.org/10.1371/journal.pone.0165367.
Mitchel, S. M., Lange, S., & Brus, H. (2013). Gendered citation patterns in international relations journals. *International Studies Perspectives*, *14*(4), 485–492. https://doi.org/10.1111/insp.12026.
Naldi, F. & Parenti, V. (2002). *Scientific and technological performance by gender. A feasibility study on patent and bibliometric indicators, II: Methodological.* https://cordis.europa.eu/pub/indicators/docs/indreportbiosoft2.pdf. Accessed 14 Feb 2017.
Naldi, F., Luzi, D., Valente, A., & Parenti, V. (2004). Scientific and technological performance by gender. In H. F. Moed, W. Glnzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research—The use of publication and patent statistics in studies of S&T systems* (pp. 299–314). Dordrecht: Kluger Academic Publishers. https://doi.org/10.1007/1-4020-2755-9.
Nielsen, M. W. (2017). Gender and citation impact in management research. *Journal of Informetrics*, *11*(4), 1213–1228. https://doi.org/10.1016/j.joi.2017.09.005.
Okon-Horodynska, E., Zachorowska-Mazurkiewicz, A., Wisla, R., & Sierotowicz, T. (2015). Gender in the creation of intellectual property of the selected European union countries. *Economics and Sociology*, *8*(2), 115–125. https://doi.org/10.14254/2071-789x.2015/8-2/9.
Paul-Hus, A., Bouvier, R. L., Ni, C. Q., Sugimoto, C. R., Pislyakov, V., & Lariviere, V. (2015). Forty years of gender disparities in Russian science: A historical bibliometric analysis. *Scientometrics*, *102*(2), 1541–1553. https://doi.org/10.1007/s11192-014-1386-4.
Podsakoff, P. M., MacKenzie, S. B., Podsakoff, N. P., & Bachrach, D. G. (2008). Scholarly influence in the field of management: A bibliometric analysis of the determinants of university and author impact in the

management literature in the past quarter century. *Journal of Management*, *34*(4), 641–720. https://doi.org/10.1177/0149206308319533.
Potthoff, M., & Zimmermann, F. (2017). Is there a gender-based fragmentation of communication science? An investigation of the reasons for the apparent gender homophily in citations. *Scientometrics*, *112*, 1047–1063. https://doi.org/10.1007/s11192-017-2392-0.
Pulkki-Brännstrom, A. M., & Stoneman, P. (2013). On the patterns and determinants of the global diffusion of new technologies. *Research Policy*, *42*(10), 1768–1779. https://doi.org/10.1016/j.respol.2013.08.011.
Qiu, J. (2008). Scientific publishing: Identity crisis. *Nature*, *451*, 766–767. https://doi.org/10.1038/451766a.
Shiller, R. J. (2017). Narrative economics. *American Economic Review*, *107*(4), 967–1004. https://doi.org/10.1257/aer.107.4.967.
Sinclair, S. & Rockwell, G. (2016). Voyant tools. Web. http://voyant-tools.org/.
Sorenson, O. (2006). Complexity, networks and knowledge flow. *Research Policy*, *35*(7), 994–1017. https://doi.org/10.1016/j.respol.2006.05.002.
Sugimoto, C. R. (2015). On the relationship between gender disparities in scholarly communication and country-level development indicators. *Science and Public Policy*, *42*(6), 789–810. https://doi.org/10.1093/scipol/scv007.
Tartari, V., & Salter, A. (2015). The engagement gap: Exploring gender differences in University-Industry collaboration activities. *Research Policy*, *44*(6), 1176–1191. https://doi.org/10.1016/j.respol.2015.01.014.
Testa, J. (2016). The Thomson reuters journal selection process. http://wokinfo.com/essays/journal-selection-process/. Accessed 18 July 2016.
Van Rijnsoever, F. J., & Hessels, L. K. (2011). Factors associated with disciplinary and interdisciplinary research collaboration. *Research policy*, *40*(3), 463–472. https://doi.org/10.1016/j.respol.2010.11.001.
Viana, M. P. (2013). On time-varying collaboration networks. *Journal of Informetrics*, *7*(2), 371–378. https://doi.org/10.1016/j.joi.2012.12.005.
Waltman, L. (2012). An empirical analysis of the use of alphabetical authorship in scientific publishing. *Journal of Informetrics*, *6*(4), 700–711. https://doi.org/10.1016/j.joi.2012.07.008.
Waltman, L., van Eck, N. J., Visser, M., & Wouters, P. (2016). The elephant in the room: The problem of quantifying productivity in evaluative scientometrics. *Journal of Informetrics*, *2*(10), 671–674. https://doi.org/10.1016/j.joi.2015.12.008.
Weichselbraun, A., Gindl, S., & Scharl, A. (2014). Enriching semantic knowledge bases for opinion mining in big data applications. *Knowledge-Based Systems*, *69*, 78–85. https://doi.org/10.1016/j.knosys.2014.04.039.
West, J., Jacquet, J., King, M., Correll, S., & Bergstrom, C. (2013). The role of gender in scholarly authorship. *PLoS ONE*, *8*, e66212. https://doi.org/10.1371/journal.pone.0066212.
Ynalvez, M. A., & Shrum, W. M. (2011). Professional networks, scientific collaboration, and publication productivity in resource-constrained research institutions in a developing country. *Research Policy*, *40*(2), 204–216. https://doi.org/10.1016/j.respol.2010.004.
Zamora-López, G., Russo, E., Gleiser, P. M., Zhou, C., & Kurths, J. (2011). Characterizing the complexity of brain and mind networks. *Philosophical Transactions of the Royal Society A*, *369*, 3730–3747. https://doi.org/10.1098/rsta.2011.0121.