Relationship among research collaboration, number of documents and number of citations: a case study in Spanish computer science production in 2000–2009

Alfonso Ibáñez · Concha Bielza · Pedro Larrañaga

Abstract This paper analyzes the relationship among research collaboration, number of documents and number of citations of computer science research activity. It analyzes the number of documents and citations and how they vary by number of authors. They are also analyzed (according to author set cardinality) under different circumstances, that is, when documents are written in different types of collaboration, when documents are published in different document types, when documents are published in different computer science subdisciplines, and, finally, when documents are published by journals with different impact factor quartiles. To investigate the above relationships, this paper analyzes the publications listed in the Web of Science and produced by active Spanish university professors between 2000 and 2009, working in the computer science field. Analyzing all documents, we show that the highest percentage of documents are published by three authors, whereas single-authored documents account for the lowest percentage. By number of citations, there is no positive association between the author cardinality and citation impact. Statistical tests show that documents written by two authors receive more citations per document and year than documents published by more authors. In contrast, results do not show statistically significant differences between documents published by two authors and one author. The research findings suggest that international collaboration results on average in publications with higher citation rates than national and institutional collaborations. We also find differences regarding citation rates between journals and conferences, across different computer science subdisciplines and journal quartiles as expected. Finally, our impression is that the collaborative level (number of authors per document) will increase in the coming years, and documents published by three or four authors will be the trend in computer science literature.

Introduction

Collaboration is a fundamental aspect of scientific research activity. Also, it is considered the key issue for solving complex problems in many areas of science (Cullen et al. 1999). Generally, scientific collaboration could be defined as researchers working together to achieve the common goal of producing new scientific knowledge.

Collaboration usually helps researchers to share their workloads, generate fresh ideas, and combine peer past experience and skills (Presser 1980; Hauptman 2005; Bammer 2008). These are all good reasons for collaboration, but they come at the expense of seeking the proper research partners, negotiating objectives, methodologies and results, managing geographic distance constraints, and communicating across organizations, cultures and disciplines and so on (Katz and Martin 1997; Landry and Amara 1998; Olson and Olson 2000; Beaver 2001).

Nowadays, researchers have begun to pay special attention to research performance and its determinants. Collaboration could be a determinant for achieving better research quality. Many researchers feel that collaborative research generally produces higher quality and more significant results than that performed by single researchers. They are motivated by the assumption that synergy leads to more and better results. A recent study explains this point by arguing that each researcher has his own knowledge and the diversity of collaborating members could be an extra resource for reinforcing research quality (Liao 2011).

Several bibliometric studies have explored the relationship of collaboration on the research performance. The relation between collaboration and productivity was first studied by Beaver and Rosen (1979). They concluded that collaboration is associated with higher productivity. Recently, Franceschet and Costantini (2010) analyzed the relationship of scholar collaboration on the impact and quality of academic papers. They noted a general positive association between the cardinality of a paper's author set and the citation impact and peer quality of the contribution. Other studies have also corroborated that research collaboration has a positive influence on the number of documents (Ponomariov and Boardman 2010) and the number of citations (Sooryamoorthy 2009).

The practice of collaboration, and especially international collaboration, is becoming a widespread phenomenon. Some studies have shown a constant increase in terms of the number of papers with international collaborations (Archibugi and Coco 2004), and an exponential increase in terms of the number of international addresses (Persson et al. 2004). This co-authorship trend is not surprising since it is an important aspect of an ideal work environment and it is also receiving interest and stimulus from policy-makers. Recent studies have analyzed the link between degree of internationalization of scientific activity and research performance at the level of individual researchers (Abramo et al. 2011a, b). They concluded that the top-performing national researchers also collaborate more abroad, but the reverse is not always true. Other studies demonstrated that the number of documents and the number of citations are positively correlated to the degree of international collaboration by a researcher (VanRaan 1998; Glanzel 2001).

It is well-known that collaboration varies across disciplines and countries. On the one hand, Gazni et al. (2012) performed a large-scale analysis to examine collaboration differences across multiple areas and from all countries. They found that the level of scientific

collaboration varies dramatically by discipline. The life sciences display high levels of co-authorship, whereas the social sciences show low levels of co-authorship. Their analysis of the collaborations between countries revealed that six countries (United States, United Kingdom, Germany, France, Italy, and Canada) account for 82 % of the world's international publications, but they are not the most collaborative countries, if measured by their proportion of collaborative output. On the other hand, Lancho-Barrantes et al. (2012) explored the provenance of the citations received by the different countries and the different types of collaborative papers. They found different percentages of papers in collaboration among countries. They also found that there is no significant correlation between scientific production and percentage of collaboration of a country. However, there is a significant negative correlation between production and the percentage traffic of citations to/from the collaborating countries. Regarding collaborative papers, they also found that there is a negative correlation between a country's production and its impact on domestic papers per paper. Finally, Franceschet and Costantini (2010) analyzed the intensity of research collaboration in different areas. They observed that collaboration is negligible in arts and humanities. They also found that the scale and formality of social science collaborations are smaller than in science disciplines. Focusing on science disciplines, collaborative work is heavily exploited in chemistry, physics, biology and medicine. In contrast, it is moderate in mathematics, engineering and computer science. Despite this, the computer science field has been expanding since 1960 in terms of both number of published papers and number of authors. Also, collaborations among different research institutes and across different countries have grown considerably recently (Franceschet 2011). According to Fortnow (2009), it is time for computer science to grow up: it is now a mature field, and no major university can survive without a strong computer science department.

Franceschet (2011) studied collaboration in computer science by means of a network science approach. Using publications from the DBLP Computer Science Bibliography, he examined properties like authors' scientific productivity and level of collaboration on papers, as well as large-scale network properties (average separation distance among scholars, distribution of the number of scholar collaborators, and dependence on star collaborators, among others). Franceschet concluded that the collaboration level in computer science papers is rather moderate (two or three authors) compared with other scientific fields. Also, he observed that the computer science collaboration network is a widely connected small world. Hence scientific information flows along collaboration links very quickly and potentially reaches almost all scholars in the discipline. Finally, he noted that the distribution of collaboration among computer science scholars is highly skewed and concentrated, where a star collaborators are responsible for a relatively high share of collaborations. Despite this, the network connectivity does not crucially depend on them. Like Franceschet (2011), we deal with bibliometric properties such as author productivity and level of collaboration on papers. Unlike Franceschet (2011), we include the number of citations and citations per document and year. Our work focuses on analyzing not network properties, but other aspects like types of collaboration, computer science subdisciplines and journal impact factor quartiles.

This paper is based on analyzing the relationship among research collaboration, number of documents and number of citations of the computer science research. Mainly, we analyze the number of documents and citations by number of authors. These measures are also analyzed (according to the author set cardinality) under different circumstances, that is, when documents are written in different types of collaboration (international, national and institutional), when documents are published in different document types (journal

article and conference paper), when documents are published in different computer science subdisciplines (artificial intelligence, cybernetics, hardware and architecture, information systems, interdisciplinary applications, software engineering and theory and methods), and, finally, when documents are published by journals with different impact factor quartiles (first-quartile journals, second-quartile journals, third-quartile journals and fourth-quartile journals). Note especially that there are no studies in the literature that investigate relationships among the above issues. Therefore, we attempt to investigate the following relationships:

– *Author cardinality versus Documents versus Citations*: We analyze the percentage evolution over time of documents published by number of authors and the average number of authors per document. We also analyze the number of citations per document and year according to the documents author set cardinality.
– *Author cardinality versus Documents versus Citations versus Types of Collaboration*: In this case, we analyzed the trend of documents published as a result of international, national and institutional collaboration by number of authors. The average number of authors per document is also analyzed according to different types of collaboration. Finally, we also explore citation measures of documents published as a result of international, national and institutional collaborations by number of authors.
– *Author cardinality versus Documents versus Citations versus Document type*: In this case, we analyzed the trend of documents published as journal articles and proceeding papers by number of authors. The average number of authors per document is also analyzed according to different document types. Finally, we also explore citation measures of documents published in journals and conferences by number of authors.
– *Author cardinality versus Documents versus Citations versus Subdisciplines*: We explore how documents published in different computer science subdisciplines change over time according to number of authors. The average number of authors per document according to different computer science subdisciplines is also analyzed. Finally, we study the number of citations per document and year in documents published in different computer science subdisciplines by author cardinality.
– *Author cardinality versus Documents versus Citations versus Impact factor*: We study the percentage trend of documents published in different journal impact factor quartiles by number of authors. The average number of authors per document according to different journal impact factor quartiles is also analyzed. We analyze citation measures of the above documents against author cardinality.

Section "Methodology" describes the data collection method. It also describes the indicators and statistical tests used in the study. The Sect. "Questions, hypotheses and results" reports the problems analyzed, our initial suppositions and results arrived regarding research collaboration. Section "Discussions and conclusions" presents final remarks and indicates possible future research directions.

Methodology

Data collection

To investigate the above relationships, this paper analyzes the publications produced by active Spanish university professors between 2000 and 2009, working in the computer science field. In the following, we illustrate the different data collection phases.

The first phase was to apply to the Spanish Ministry of Education for a list of academics associated with the computer science area who were active as of December 31, 2009. This list includes the full name of 2004 academics, and their associated university, position and research area. These researchers are attached to the main area in which they lecture and regularly publish by the Spanish Ministry of Education

The next phase was to retrieve a list of publications and citation data (from January 1, 2000 to December 31, 2009) for each academic. This information was carefully downloaded from the Web of Science (Web of Knowledge) bearing in mind Spanish personal name variations in international databases (Ruiz-Perez et al. 2002). After that, only documents considered as journal articles and conference papers were taken into consideration. Also, we used the publication subject classification as a filter. In this way, we only selected documents which were published in journals and conferences belonging to the seven major fields of computer science. According to the Journal Citation Reports these major fields are: artificial intelligence, cybernetics, hardware and architecture, information systems, interdisciplinary applications, software engineering and theory and methods. The result was around 20,000 publications. Finally, we manually checked those publications in which only one affiliation is available as reported by Web of Science against the original published publications. We noted that Web of Science only provides the affiliation of the first author for some publications (especially conference papers). In order to ensure the reliability of results, we checked our final list of publications against other databases like DBLP Computer Science Bibliography, personal webpages and institutional websites, among others. Also, the impact factors of journals belonging to each of the seven major fields of computer science were extracted from the corresponding Journal Citation Reports edition (2000–2009).

The last phase was to develop software which used all this information in order to calculate some indicators (Section "Indicators") by number of authors. This dataset was also used in Ibañez et al. (2012) to characterize research activity of Spanish universities and their academic staff, identifying both their strengths and weaknesses nationwide. The analysis carried out is also performed by autonomous regions, public universities, subject areas and professional standing.

Indicators

The number of documents and citations are indispensable for analyzing research activity. Citations are measures of information use, reception and, in a way, of influence (Cronin 1981). They can be considered as an indirect measure of publications quality in most cases, although there may be retracted papers that receive a lot of citations.

We also computed two measures of collaboration which are generally used in studies of research collaboration (Levitt and Thelwall 2009). These measures are the collaborative rate and the collaborative level. Collaborative rate (CR) is the percentage of documents with more than one author, whereas collaborative level (CL) is the average number of authors per document. These measures are computed by analyzing the number of authors of each publication.

Regarding the measures of internationalization (Abramo et al. 2011b), we use the international rate (IR) to analyze the percentage of papers that have been produced in collaboration with foreign institutions, that is, the percentage of publications co-authored with at least one co-author from an foreign institution. This measure is computed by analyzing the publications whose affiliations include addresses from more than one country.

Finally, the impact factor (IF) defines the status of a journal for a specific year as the number of citations received in that year over the number of articles published in the journal in the two previous years. It is still recognized as the primary measure of journal quality and has a major influence on scientific behavior (Weingart 2005). Furthermore, experience has shown that the best journals in each specialty are the publications in which it is hardest to get an article accepted, and these are the journals that have a high impact factor (Garfield 2000).

Statistical tests

Statistical tests determine whether there is enough evidence to reject a conjecture about the data. The conjecture is called the null hypothesis. Not rejecting the conjecture may be a good result if we want to continue to act as if we believe the null hypothesis is true. Or it may be a disappointing result, possibly indicating that we may not yet have enough information to reject the null hypothesis.

Tests that do not make assumptions about the population distribution are referred to as non-parametric tests. All commonly used non-parametric tests rank the outcome variable from low to high and then analyze the ranks.

In this paper, we use two non-parametric tests: Kruskal-Wallis (1952) test and Mann-Whitney (1947) test. The Kruskal-Wallis test analyzes whether three or more samples could have come from the same distribution. The null hypothesis is that the populations from which the samples originate have the same distribution. When the Kruskal-Wallis test leads to significant results, then at least one of the samples is different from the other samples. The test does not identify where the differences occur or how many differences actually occur. In contrast, the Mann-Whitney test analyzes whether two samples could have come from the same distribution. It is helpful for analyzing the specific sample pairs for significant differences. The significance level of these tests was 0.05 in all cases.

Questions, hypotheses and results

This paper analyzes the relationship among collaboration (number of authors), documents and citations on several issues such as types of collaboration, document types, computer science sub-disciplines and journal impact factor quartiles. The number of authors has been grouped into six different subsets (1 author, 2 authors, 3 authors, 4 authors, 5 author and >5 authors).

How do productivity and utility vary according to author cardinality?

The first question investigates the number of documents and citations of documents published by number of authors. Our first impression is that computer science documents are usually the result of collaboration. Specially, we believe that the average document is written by three or four authors. This is based on the idea that different co-authors reinforce research quality. We also think that the number of authors per document has gradually increased in the last decade. Regarding utility, we believe that a greater number of authors can lead to a higher number of citations because co-authors are more likely to disseminate the document.

According to the different author subsets, document distribution in the analyzed period was: 1 author (2.651 %), 2 authors (18.182 %), 3 authors (33.037 %), 4 authors
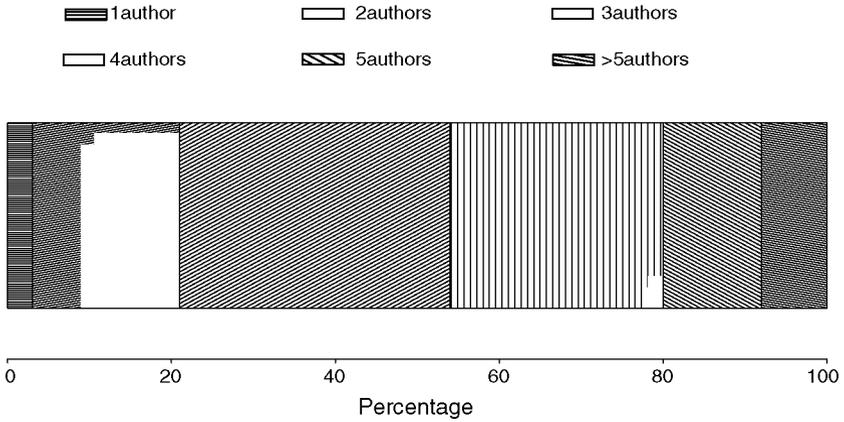
Fig. 1 Percentage of published documents by number of authors
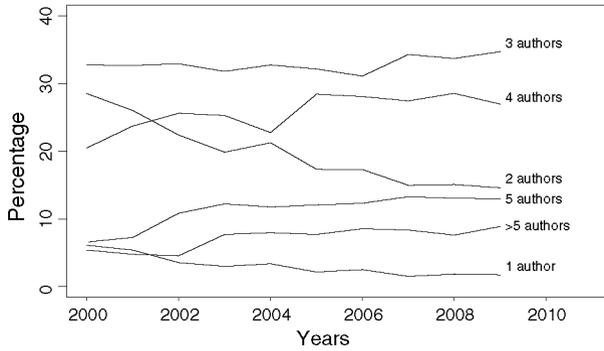
Fig. 2 Evolution of percentage of published documents by number of authors

(26.456 %), 5 authors (11.994 %), and >5 authors (7.680 %). We found that most docu-
ments were published by three and four authors, whereas single-authored documents
accounted for the lowest percentage (see Fig. 1).

Figure 2 plots the evolution of published documents by number of authors from 2000 to
2009. Analyzing the number of authors, Fig. 2 shows that the percentage of documents
published by different authors underwent some changes in the last decade. In earlier years,
documents published by two authors accounted for a sizeable percentage of total publi-
cations (28.538 % in 2000), but in 2009, it represented 14.616 % of total publications. In
contrast, the percentage of documents with three or more authors increased. The per-
centage of documents published by one author also decreased over the analyzed years, and,
therefore, the collaborative rate increased over time.

As expected, these results bear out previous works stating that the practice of collab-
oration is becoming a widespread phenomenon. The number of authors used to be lower
than it is today. Just a few authors were responsible for the hypothesis, experimental
design, results and conclusion (Zetterstrom 2004). Nowadays, most projects require the
participation of many researchers, who are all entitled to be authors when the results are
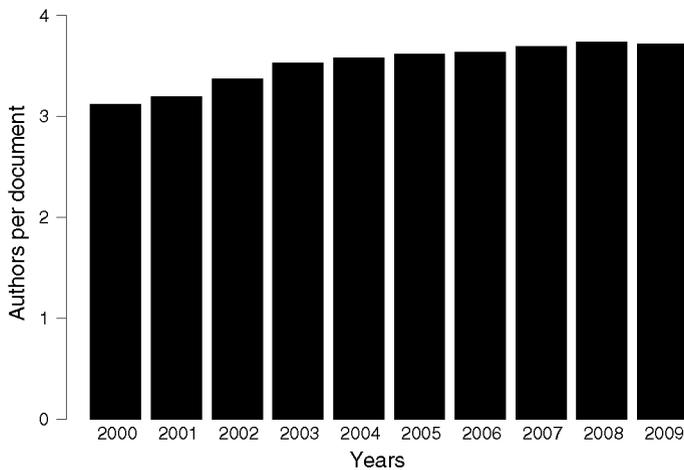
Fig. 3 Evolution of the average number of authors per document

reported. Other reasons that have increased the number of authors per document in recent years are dependency on the department chair and the addition of influential authors to raise a paper's prestige, among others. These authors who are neither author neither contributors are called guests (Laine and Mulrow 2005) or even parasites (Solomon 2009).

Figure 3 analyses collaboration level. It shows the evolution of the average number of authors per document. Collaborative level has increased in the last few years. Values rose from 3.118 authors in 2000 to 3.739 authors in 2008, so the increase was 19.917 %. Taking the 2009 year as an example publication year, we observed that the published documents had an average of 3.721 authors per document. Analyzing these values, our impression is that documents published by three or four authors will be the trend in computer science literature in the coming years.

Table 1 shows the average number of citations, the average number of years since the publication year, and the average number of citations per document and year (columns). These measures and their standard deviations are calculated for each different number of authors (rows).

Analyzing Table 1, we observed that the highest average number of citations (3.019 ± 7.260) corresponded to documents published by one author, whereas the lowest average value (1.852 ± 4.830) corresponded to documents published by five authors (column 2). These results were influenced by publication age, that is, the number of years since the publication year. We accounted for this point by calculating the average age (column 3). We observed that documents published by one author had the highest average age (5.536 ± 2.821), whereas documents published by more than five authors had the lowest average age (4.172 ± 2.460). We calculated the number of citations per document and year as an accurate measure for comparing documents published by number of authors. This ratio, which is a utility measure is a possible indirect measure of the document's quality. Analyzing this measure, we found that documents published by two authors had the highest value (0.478 ± 1.293), and documents published by five authors had the lowest value (0.363 ± 0.841). Note that documents published by one or two authors had higher values of citations per document and year than documents published by three or more authors. A possible explanation could be that an important percentage of documents

Table 1 Mean ± standard deviation of citation measures for documents published by different number of authors

Authors	Citation count	Publication age	Citations ratio
1	3.019 ± 7.260	5.536 ± 2.821	0.443 ± 0.889
2	2.943 ± 10.297	5.023 ± 2.733	**0.478 ± 1.293**
3	2.181 ± 6.562	4.450 ± 2.634	0.407 ± 1.021[†]
4	1.882 ± 5.469	4.346 ± 2.548	0.365 ± 0.903[†]
5	1.852 ± 4.830	4.200 ± 2.453	0.363 ± 0.841[†]
>5	1.913 ± 5.913	4.172 ± 2.460	0.409 ± 1.090[†]

[†] Represents those results that are statistically different in citations ratio with respect to the benchmark subset (highlighted in boldface)

published by one or two authors are review papers. A review paper is usually written by a single senior research, and it is likely to be cited extensively. This would explain why single-authored documents received more citations than documents written by a larger number of researchers.

The results of the Kruskal-Wallis test showed that there were significant differences among the six author subsets on the basis of average number of citations per document and year. So, we run Mann-Whitney tests in order to find out which subsets rank better according to this criterion. We compared documents published by two authors (benchmark subset), which had the highest average value, with the other documents. Subsets marked in Table 1 with the symbol † had statistically significant differences with respect to the benchmark subset (highlighted in boldface). Results show that there were significant differences between the 2-author subset and subsets with more authors. In contrast, results do not show statistically significant differences between the 2-author subset and 1-author subset. Unlike Franceschet and Costantini (2010), we did not find a positive association between the author set cardinality of a document and citation impact.

How do productivity and utility in different types of collaboration vary according to author cardinality?

The second question analyzes whether productivity and utility behave differently across different types of collaboration. We make a distinction between three types of collaboration: international, national and institutional cooperation. International collaboration refers to co-authorship by researchers from both national and foreign institutions. National collaboration refers to co-authorship by researchers belonging to different institutions in the same country. Finally, institutional collaboration refers to co-authorship among researchers belonging to the same institution.

Due to problems of geographic distance and communication across organizations, we believe that most documents are written through institutional collaboration. We also believe that there will be more authors per document resulting from international collaboration than via national and institutional collaboration. On the other hand, analyzing utility across different types of collaboration, it is reasonable to expect, precisely because of the differences among authors, the quality of documents resulting from international collaborations to be greater, and have a higher number of citations. Also, we believe that a greater number of authors can lead to a higher number of citations for a particular type of collaboration.

The document distribution in the analyzed period was: international collaboration (13.334 %), national collaboration (13.112 %) and institutional collaboration (73.554 %). So, the value of the international rate (IR) was 13.334 %. We found that this percentage was very similar on a year-by-year basis. Therefore, the evolution of the international rate has not undergone major changes in the analyzed period. We also found that most collaborative documents were published via institutional collaboration.

Figure 4 represents the evolution of published documents by number of authors and type of collaboration. Regarding international collaboration, Fig. 4a shows that the percentage of documents published by number of authors has recently undergone changes. We found that documents published by three and four authors represented the highest percentages of published documents each year, whereas documents published by two, five, and more than five authors represented the lowest percentages. Analyzing Fig. 4a, we found a sizeable decrease in the percentages associated with documents published by two and four authors (e.g., the percentage of documents published by four authors was 39.130 % in 2000 and 24.476 % in 2009). In contrast, the number of documents published by three authors fluctuated considerably, and there were increases in the number of documents published by five or more authors. Regarding these increases, we show that percentages associated with documents with five authors rose from 6.522 % in 2000 to 16.776 % in 2008, and percentages associated with documents with more than five authors rose from 10.526 % in 2001 to 18.182 % in 2009.

By national collaboration (see Fig. 4b), results show an important decrease of documents published by two authors. The percentages associated with these documents were 19.022 % in 2004 and 7.189 % in 2009. Likewise, the percentage of documents with three authors also decreased from 41.818 % in 2000 to 30.719 % in 2009. In contrast, percentages associated with documents published by four or more authors increased over the time period.

Figure 4c analyzes institutional collaboration. We observed a sharp decrease in the percentage of documents published by two authors. In earlier years, collaboration between two authors represented a sizeable percentage (35.353 %) of the total publications, but this percentage decreased considerably (16.101 %) in 2009. The other percentages associated with documents with three or more authors have gradually increased over last few years.

Finally, note that publication behavior has been similar across different types of collaboration in recent years. There were two different groups: documents published by three or four authors which had the highest percentages, and documents published by two, five, or more than five authors that had the lowest percentages. These groups were also highlighted in Fig. 2 plotting the evolution of the percentage of published documents by number of authors. Figure 2 also shows an important decrease of documents published by two authors via international, national and institutional collaborations.

Regarding collaborative level, Fig. 5 shows the average number of authors per document for each type of collaboration. Taking 2009 as an example publication year, we observed that the average international document was published by 4.273 authors, whereas the average national document was published by 4.111 authors, and the average institutional document was published by 3.603 authors.

According to the above values and the evolution illustrated in Fig. 5, we found that international collaborations usually had the highest number of authors per document, followed by national and institutional collaborations, as expected. A large number of international and national collaborations spring from projects that require the participation of many researchers from different institutions, whereas most institutional collaborations usually involve authors from the same research group. For these reasons, both international
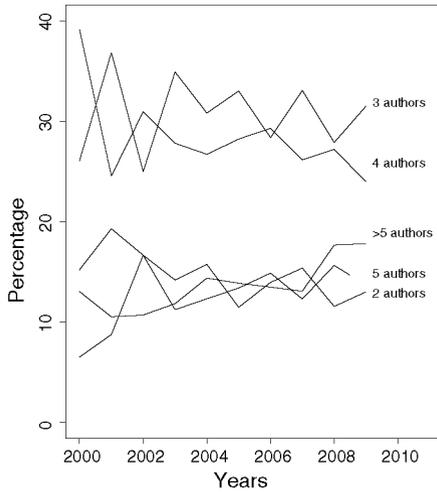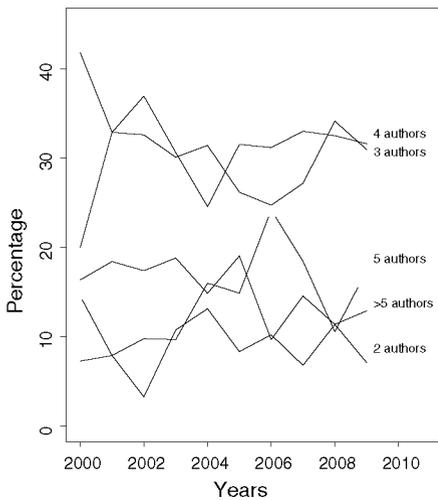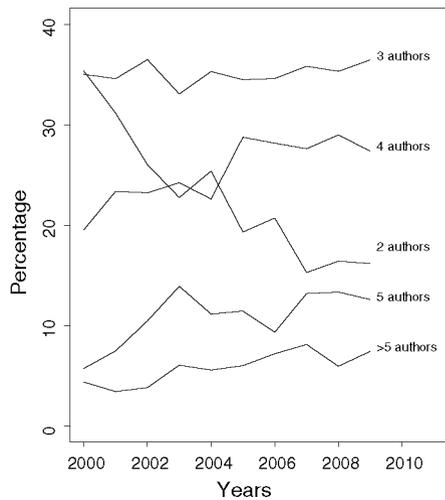
Fig. 4 Evolution of percentage of published documents by number of authors and type of collaboration (international, national and institutional)

and national collaborations involve more authors than institutional collaborations, increasing the number of authors per document.

Table 2 shows the average number of citations, the average number of years since the publication year, and the average number of citations per document and year of documents published via different types of collaboration (international, national and institutional) and written by different numbers of authors. It also shows the standard deviations associated with the above measures.

We found that international collaborations usually had the highest average values of citations per document and year for different numbers of authors, followed by national
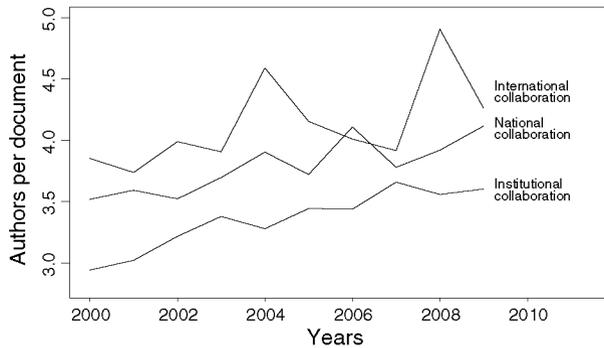
Fig. 5 Evolution of the average number of authors per document according to different types of collaboration (international, national and institutional)

collaborations and institutional collaborations. International collaboration often involves more authors than other types of collaboration as mentioned before. As the authors are likely to disseminate the document, it is reasonable to assume that there will be a greater number of citations. Taking documents published by more than five authors as an example, note that the average values of international, national and institutional documents were 0.837 ± 1.816, 0.506 ± 0.804 and 0.207 ± 0.607, respectively. Like VanRaan (1998) and Glanzel (2001), our results demonstrate that, on an average, international collaboration results in documents with higher citation rates than national and institutional documents.

A Kruskal-Wallis test was performed in order to compare different subsets of authors (according to a particular type of collaboration). The Kruskal-Wallis test did not find statistically significant differences across international and national documents published by different authors. In contrast, results show that there were significant differences among institutional documents published by different authors. So, several Mann-Whitney tests were carried out to find out which subsets of authors (highlighted by †) were significantly different from the benchmark subset (highlight in boldface). We found that institutional documents published by two authors were significantly different to all other subsets of authors. Analyzing the statistical test results, we conclude that it is better to publish with few authors in order to improve document utility at the institutional level, whereas the number of authors does not affect the average number of citations per document and year at the national and international level.

How do productivity and utility in different document types vary according to author cardinality?

The third question analyzes whether productivity and utility behave differently across different document types. Journal articles and conference papers are the document types studied in this paper.

We believe that publication behavior is different across journals and conferences. Due to the undeniable advantage of conferences (provide fast and regular publication of papers and bring researchers together by offering the opportunity to present and discuss the paper with peers), we think that authors tend to publish more documents in conferences than in journals. We also think that most journal articles and conferences papers are published by three or four authors. By collaborative level, we suppose that there are no clear differences

Table 2 Mean ± standard deviation for citation measures of international, national and institutional collaborations in documents published by number of authors

	Collaborations		
	International	National	Institutional
2-authors			
Citation count	4.899 ± 12.786	5.216 ± 18.433	2.381 ± 8.127
Publication age	4.974 ± 2.487	5.584 ± 2.384	4.955 ± 2.800
Citations ratio	0.831 ± 1.790	**0.749 ± 2.066**	**0.395 ± 1.046**
3-authors			
Citation count	3.765 ± 6.813	3.995 ± 8.608	1.620 ± 6.022
Publication age	4.799 ± 2.413	5.131 ± 2.669	4.281 ± 2.642
Citations ratio	0.716 ± 1.182	0.722 ± 1.218	0.305 ± 0.932[†]
4-authors			
Citation count	4.320 ± 8.628	3.002 ± 6.267	1.141 ± 4.104
Publication age	4.909 ± 2.453	4.931 ± 2.534	4.104 ± 2.532
Citations ratio	0.787 ± 1.453	0.599 ± 1.093	0.228 ± 0.634[†]
5-authors			
Citation count	3.552 ± 6.858	2.682 ± 5.201	1.285 ± 4.028
Publication age	4.551 ± 2.331	4.621 ± 2.383	4.021 ± 2.478
Citations ratio	0.671 ± 1.146	0.573 ± 1.047	0.245 ± 0.662[†]
>5-authors			
Citation count	3.750 ± 9.821	2.579 ± 4.078	0.980 ± 3.588
Publication age	4.489 ± 2.456	4.794 ± 2.483	3.869 ± 2.412
Citations ratio	**0.837 ± 1.816**	0.506 ± 0.804	0.207 ± 0.607[†]

[†] Represents those results that are statistically different in citations ratio with respect to the benchmark subset (highlighted in boldface)

between journals and conferences. On the other hand, we believe that citation counts received by journal articles are higher than received by conference papers because of their prestige. Furthermore, we also think that multi-authored documents receive more citations than single-authored documents.

The document distribution in the analyzed period was: journal articles (32.262 %) and conference papers (67.738 %). These percentages bear out previous works, like Franceschet (2010), stating that 1/3 of computer science literature are journal articles and 2/3 are conference papers. On a year-by-year basis the percentage of journal articles vary from 26.000 to 44.792 %, whereas conference papers vary from 55.208 to 74.000 %. We also found that the percentage of conference papers have gradually decreased. In 2005, conference papers accounted for a sizeable percentage of total publications (74.000 %), but in 2009, it represented 62.678 % of total publications. An interpretation could be that researchers are progressively shifting from conferences to journals, considering budget shortages or higher prestige of journals over conferences.

According to the number of authors, we found that 54.098 % of single-authored documents are published in journals, whereas 45.902 % are published in conferences. The rest of percentages were: 2 authors (43.098 % in journals and 56.902 % in conferences), 3 authors (36.906 % in journals and 63.094 % in conferences), 4 authors (32.839 % in

journals and 67.161 % in conferences), 5 authors (31.750 % in journals and 68.250 % in conferences), >5 authors (34.002 % in journals and 65.998 % in conferences).

Figure 6 shows the evolution of the percentage of documents published in computer science journals and conferences by number of authors from 2000 to 2009. We found that the percentage of documents associated with each author subset was similar in journal articles and conference papers, so there are no important differences in publication behavior by number of authors between journals and conferences. In general, we observed that there was a decrease in the number of documents published by one and two authors in both cases. We also observed that documents written by three and four authors accounted for the highest percentages, whereas the lowest percentage of documents were written by one author. Taking the journal articles as an example, Fig. 6a shows that the percentage of documents with four or more authors has gradually increased over the last few years. Specially, documents published by four authors have undergone an increase in the last few years, they accounted for 18.116 % of all publications in 2004, and 27.687 % of total publications in 2009. In contrast, documents published by one author and two authors have decreased over the analyzed years and we noted that single-authored documents account for the lowest percentage in the 2002–2009 period.

Regarding collaborative level, Fig. 7 shows the average number of authors per document for journal articles and conference papers. According to its evolution, conference papers have had the highest number of authors per document in earlier years. Despite this, journal articles and conference papers had similar number of authors per document in recent years. Taking 2009 as an example publication year, we observed that the average journal article was published by 3.738 authors, whereas the average conference paper was published by 3.711 authors.

Table 3 shows the average number of citations, the average number of years since the publication year, and the average number of citations per document and year. These
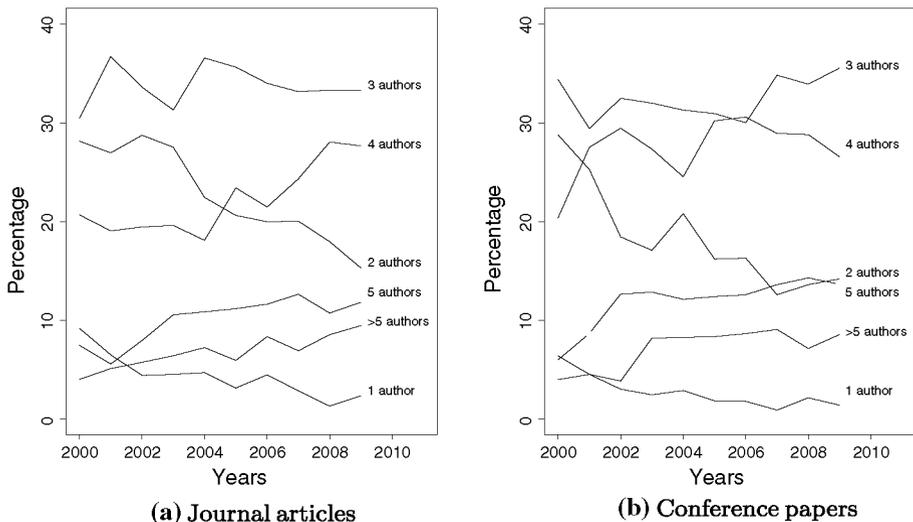

Fig. 6 Evolution of percentage of published documents by number of authors and document type
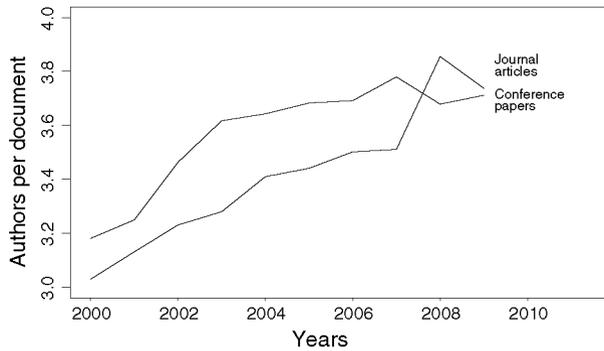
Fig. 7 Evolution of the average number of authors per document according to different document types

Table 3 Mean ± standard deviation for citation measures of journal and conference documents published by number of authors

	Document type	
	Journal article	Conference paper
1-author		
Citation count	4.983 ± 9.991	1.371 ± 2.757
Publication age	5.783 ± 2.937	5.329 ± 2.713
Citations ratio	0.698 ± 1.171	**0.229 ± 0.456**
2-authors		
Citation count	6.029 ± 15.774	1.063 ± 3.133
Publication age	5.092 ± 2.858	4.981 ± 2.654
Citations ratio	0.940 ± 1.917	0.196 ± 0.495
3-authors		
Citation count	5.043 ± 10.547	0.770 ± 1.865
Publication age	4.534 ± 2.794	4.408 ± 2.551
Citations ratio	0.923 ± 1.581	$0.153 ± 0.358^{\dagger}$
4-authors		
Citation count	4.753 ± 9.045	0.746 ± 2.205
Publication age	4.142 ± 2.776	4.427 ± 2.448
Citations ratio	0.924 ± 1.457	$0.143 ± 0.354^{\dagger}$
5-authors		
Citation count	4.702 ± 7.457	0.717 ± 2.449
Publication age	4.118 ± 2.603	4.233 ± 2.392
Citations ratio	0.919 ± 1.251	$0.142 ± 0.442^{\dagger}$
>5-authors		
Citation count	4.481 ± 9.601	0.783 ± 2.388
Publication age	3.974 ± 2.657	4.259 ± 2.366
Citations ratio	**0.971 ± 1.734**	$0.161 ± 0.434^{\dagger}$

† Represents those results that are statistically different in citations ratio with respect to the benchmark subset (highlighted in boldface)

measures and their standard deviations are calculated for each different number of authors and document type.

Analyzing the number of authors in Table 3, we noted that documents published by more than five authors had the highest average value of citations per document and year (0.971 ± 1.734) when they were published by journals. In contrast, single-authored documents had the highest average value of citations per document and year (0.229 ± 0.456) when they were published by conferences. As expected, journal articles had higher citations per document and year than conference papers. These results corroborate previous work like Franceschet (2010), in which the impact of journal publications was significantly higher than the impact of conference papers.

We performed a Kruskal-Wallis test in order to compare subsets of different authors across documents published in journals and conferences. Results show that there were no significant differences across documents published by journals. In contrast, it found significant differences across documents published by conferences: the average number of citations per document and year of documents published by one author (0.229 ± 0.456) was significant different (higher) to documents published by three authors (0.153 ± 0.358), four authors (0.143 ± 0.354), five authors (0.142 ± 0.442) and more than five authors (0.161 ± 0.434).

How do productivity and utility in different computer science subdisciplines vary according to author cardinality?

The fourth question investigates the productivity and utility of authors across the seven computer science subdisciplines: artificial intelligence, cybernetics, hardware and architecture, information systems, interdisciplinary applications, software engineering and theory and methods.

We believe that publication behavior is different across subdisciplines. We think that authors tend to publish more documents in mature disciplines like theory and methods. Also, we believe that the percentages of documents published by a specific number of authors are similar across subdisciplines. We think that most documents are published by three or four authors in all subdisciplines. Despite this, we believe that the collaborative level is different. We think that interdisciplinary applications documents are usually written by more authors than publications in other disciplines. This idea is based on the assumption that interdisciplinary applications documents could be published by authors belonging to many different areas, resulting in more authors per document. By utility, we also believe that citation counts are different across subdisciplines. We think that a greater number of authors leads to a higher number of citations in any particular a subdiscipline.

We found that according to the Web of Knowledge there is an overlap across the seven subdisciplines. Thus, one document could belong to more than one discipline at the same time. The document distribution in the analyzed period was: artificial intelligence (24.849 %), cybernetics (1.613 %), hardware and architecture (7.285 %), information systems (9.528 %), interdisciplinary applications (5.543 %), software engineering (11.059 %) and theory and methods (40.123 %). We found that most documents were related to theory and methods, whereas cybernetics accounted for the lowest percentage of published documents.

Figure 8 shows the evolution of the percentage of documents published in computer science subdisciplines by number of authors from 2000 to 2009. After analyzing all computer science subdisciplines in Fig. 8, we found that the percentage of documents associated with each author subset was quite alike across different subdisciplines. These

percentages were: 1 author (2.5274–3.068 %), 2 authors (18.001–22.085 %), 3 authors (32.594–33.247 %), 4 authors (24.773–26.238 %), 5 authors (9.811–11.967 %) and >5 authors (7.191–8.333 %). According to these percentages, we found that there were no important differences in publication behavior by number of authors across subdisciplines. Looking at all the charts illustrated in Fig. 8, we also observed similarities across subdisciplines. We found that there was a general decrease in the number of documents published by one and two authors in all subdisciplines. Also, we observed that documents written by three and four authors accounted for the highest percentage in all subdisciplines, whereas the lowest percentage of documents were written by one author. On the other hand, the percentages associated with each subdiscipline have fluctuated widely in most computer science subdisciplines in the last decade. In contrast, artificial intelligence (see Fig. 8a) and theory and methods (see Fig. 8g) did not experience as many fluctuations as other subdisciplines. We found that these two subdisciplines behaved very like computer science generally (see Fig. 2). This was reasonable because these subdisciplines had the highest percentages of published documents, 24.849 and 40.123 %, respectively.

Taking the artificial intelligence discipline as an example, Fig. 8a shows that documents published by two authors accounted for the highest percentage (34.759 %) of all publications in 2000, but represented only 16.145 % of total publications in 2009. Documents published by one author have also decreased over the analyzed years and accounted for the lowest percentage in the 2002–2009 period. In contrast, the percentage of documents with three or more authors has gradually increased over the last few years.

Figure 9 analyses collaborative level. It shows the evolution of the average number of authors per document according to different subdisciplines. These measures have tended to increase over the last few years. We emphasize the hardware and architecture subdiscipline whose values rose from 3.162 in 2000 to 4.405 in 2009. It was an increase of 39.311 % In contrast, we found that the number of authors per cybernetics document underwent a sizeable decrease up until 2004 (27.248 %), and later recovered. We also found that the range of the average number of authors per document was different across subdisciplines with respect to the analyzed year. Despite this, we found that the range was wider in earlier years (2000–2004) than in later years (2005–2009). Finally, we found that the highest values for collaborative level were achieved by documents belonging to the interdisciplinary applications (4.405 authors per document) and hardware and architecture (4.074 authors per document) subdisciplines in 2009. These values were the result of a major increment of documents published by more than three authors in these subdisciplines over the last few years (see Fig. 8).

Table 4 shows the average number of citations, the average number of years since the publication year, and the average number of citations per document and year of documents published in different subdisciplines and written by different numbers of authors. It also shows standard deviations of the above measures. Analyzing the average number of citations per document and year, we observed that some subdisciplines were more often cited than others. It is noteworthy that artificial intelligence documents, which had a lower value of authors per document than other subdisciplines, usually had a higher average values of citations per document and year than others. In contrast, hardware and architecture documents, which had the highest collaborative level value in recent years, received fewer citations than other subdisciplines like artificial intelligence, cybernetics and information systems. Citation counts by subdisciplines were known to vary within a particular discipline (Bornmann and Daniel 2006). Some studies, like Smolinsky and Lercher (2012), found that citation practices differ across mathematics subdisciplines. Like
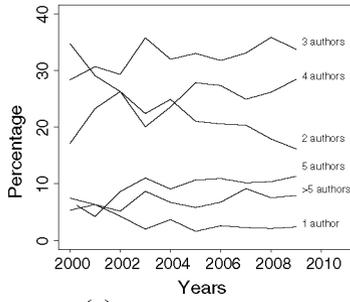
(a) AI:Artificial Intelligence
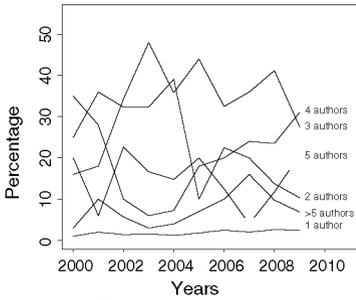
(b) CB: Cybernetics
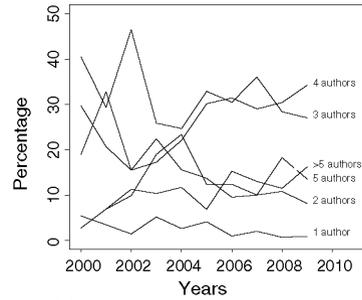
(c) HA: Hardware and Architecture
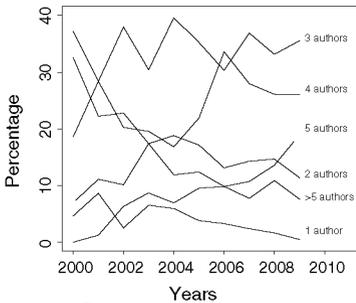
(d) IS: Information Systems
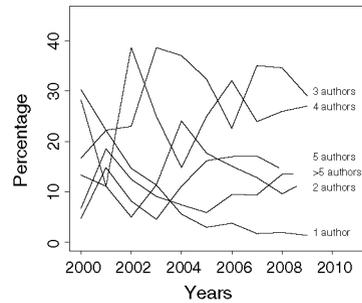
(e) IA :Interdisciplinary Applications
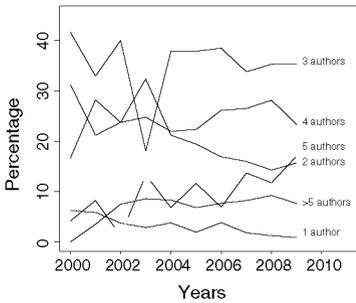
(f) SE: Software Engineering
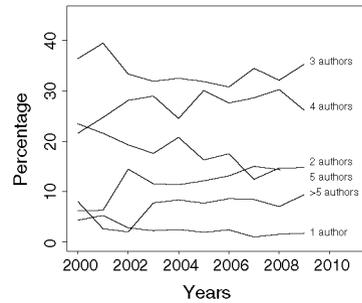
(g) TM: Theory and Methods

Fig. 8 Evolution of percentage of documents published in computer science subdisciplines by number of authors
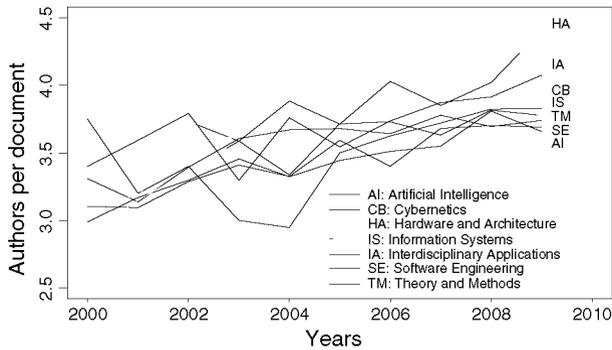
Fig. 9 Evolution of the average number of authors per document by disciplines

Smolinsky and Lercher (2012), we also found different citation behaviors by subdisciplines within a specific discipline (computer science in our case).

In order to compare citation behaviors by author subsets for a particular subdiscipline, several Kruskal-Wallis tests were performed. The Kruskal-Wallis test did not find meaningful differences across documents belonging to information systems, interdisciplinary applications and software engineering. In contrast, results show that there were significant differences among documents belonging to artificial intelligence, cybernetics, hardware and architecture, and theory and methods (see numbers in boldface and the † symbols). Taking artificial intelligence documents as an example, Table 4 shows that the number of citations per document and year of documents published by two authors (0.663 ± 1.736) were significantly different to documents published by three (0.515 ± 1.323) and four (0.551 ± 1.283) authors. Similarly, we found that the number of citations per document and year of hardware and architecture documents published by two authors (0.435 ± 1.033) were significantly different to documents published by three (0.334 ± 0.996) and four authors (0.229 ± 0.641).

Analyzing the statistical test results in Table 4, we conclude that the number of authors does not always affect the average number of citations per document and year. In contrast, specific subdisciplines, like artificial intelligence, cybernetics, hardware and architecture, and theory and methods, are affected by the number of authors. Unlike Franceschet and Costantini (2010) who found a general positive association between a paper's author set cardinality the citation impact, we observe that documents with fewer authors usually have the highest average value of citations per document and year. Specifically, documents written by one author have the highest values for information systems, software engineering and theory and methods, whereas documents written by two authors have the highest values in artificial intelligence, cybernetics and hardware and architecture. In contrast, we note that interdisciplinary applications documents published by more than five authors have the highest number of citations per document and year.

How do productivity and utility in different journal impact factor quartiles vary according to author cardinality?

Journals ordered by impact factor can be organized into four quartiles. The first quartile denotes the top 25 % of the impact factor distribution, the second quartile means a middle-high position (between the top 50 % and top 25 %), the third quartile is a middle-low

Table 4 Mean ± standard deviation for citation measures of documents published by numbers of authors according to seven subdisciplines

	AI	CB	HA	IS	IA	SE	TM
1-author							
Citation count	2.490 ± 5.544	0.500 ± 0.707	1.778 ± 2.290	5.026 ± 12.946	2.348 ± 3.725	4.400 ± 12.740	2.514 ± 4.219
Publication age	5.792 ± 3.067	2.500 ± 2.121	5.944 ± 2.689	5.949 ± 2.502	5.632 ± 2.790	5.714 ± 2.771	5.180 ± 2.728
Citations ratio	0.360 ± 0.727	0.125 ± 0.177	0.295 ± 0.403	**0.625 ± 1.483**	0.403 ± 0.581	**0.557 ± 1.383**	**0.427 ± 0.653**
2-authors							
Citation count	4.263 ± 13.641	7.952 ± 21.971	2.838 ± 8.624	2.335 ± 7.479	1.826 ± 4.359	1.649 ± 4.062	2.451 ± 9.990
Publication age	5.292 ± 2.825	5.357 ± 2.458	5.581 ± 2.847	5.031 ± 2.899	4.174 ± 2.510	4.826 ± 2.742	4.735 ± 2.561
Citations ratio	**0.663 ± 1.736**	**1.059 ± 2.486**	**0.435 ± 1.033**	0.479 ± 1.272	0.363 ± 0.747	0.291 ± 0.639	0.394 ± 1.184
3-authors							
Citation count	2.798 ± 8.209	4.242 ± 9.980	1.878 ± 5.959	1.817 ± 4.674	1.510 ± 3.372	1.844 ± 4.841	1.793 ± 5.940
Publication age	4.622 ± 2.702	4.323 ± 2.289	4.776 ± 2.793	4.301 ± 2.722	3.784 ± 2.451	4.221 ± 2.694	4.397 ± 2.549
Citations ratio	0.515 ± 1.323†	0.763 ± 1.359	0.344 ± 0.996†	0.386 ± 1.038	0.318 ± 0.670	0.375 ± 0.908	0.334 ± 0.881†
4-authors							
Citation count	2.743 ± 7.271	2.657 ± 5.886	1.307 ± 4.694	2.108 ± 6.930	1.476 ± 4.821	1.754 ± 4.484	1.337 ± 3.871
Publication age	4.494 ± 2.660	3.714 ± 2.729	4.101 ± 2.655	4.101 ± 2.670	3.693 ± 2.515	4.196 ± 2.683	4.334 ± 2.426
Citations ratio	0.551 ± 1.283†	0.731 ± 1.802	0.229 ± 0.641†	0.399 ± 0.995	0.338 ± 0.777	0.338 ± 0.783	0.258 ± 0.639†
5-authors							
Citation count	3.046 ± 6.821	3.885 ± 9.868	2.130 ± 5.459	1.346 ± 3.906	2.489 ± 5.681	1.293 ± 3.609	1.377 ± 4.107
Publication age	4.375 ± 2.571	4.500 ± 2.746	4.200 ± 2.436	3.949 ± 2.716	3.298 ± 2.233	3.487 ± 2.432	4.120 ± 2.327
Citations ratio	0.593 ± 1.167	0.548 ± 1.087	0.384 ± 0.820	0.220 ± 0.593	0.579 ± 1.166	0.292 ± 0.729	0.272 ± 0.694†
>5-authors							
Citation count	2.364 ± 5.534	2.143 ± 5.405	2.206 ± 6.390	1.478 ± 3.510	2.451 ± 5.995	1.388 ± 3.612	1.626 ± 6.819
Publication age	4.467 ± 2.669	3.643 ± 2.307	3.979 ± 2.504	3.696 ± 2.208	3.549 ± 2.666	3.825 ± 2.341	4.069 ± 2.354
Citations ratio	0.509 ± 1.320	0.277 ± 0.662†	0.434 ± 1.055	0.396 ± 0.939	**0.776 ± 2.090**	0.318 ± 0.736	0.321 ± 0.998†

† Represents those results that are statistically different in citations ratio with respect to the benchmark subset (highlighted in boldface)

position (top 75 % to top 50 %), and the fourth quartile represents a bottom position (bottom 25 % of the impact factor distribution). The fourth question investigates the number of authors across different quartiles. Also, we analyze the productivity and utility of documents published in different journal impact factor quartiles according to the author cardinality.

We have the impression that first-quartile journals have the lowest publication rate due to their selective strategy, that is, low acceptance rates. Regarding the number of authors, we believe that the percentages of documents published by a specific number of authors are similar across quartiles. We think that most documents are published by three or four authors in each quartile. By collaborative level, we suppose that there are no clear differences across different quartiles. So, we also think that three or four authors per document is the average collaborative level value. On the other hand, we believe that citation counts are obviously different across quartiles. Furthermore, we also think that multi-authored documents receive more citations than single-authored documents within a specific quartile.

Table 5 shows the percentages of documents published in each quartile for different numbers of authors. In single-authored documents the percentages associated with each quartile were: first-quartile (28.333 %), second-quartile (24.167 %), third-quartile (31.667 %) and fourth-quartile (15.833 %). In this case, the third-quartile had the highest percentage of published documents. This quartile also had the highest percentage of documents published by two authors. In contrast, documents published by three or more authors were usually published in journals belonging to the first quartile. On the other hand, we found that journals belonging to the fourth quartile accounted for the lowest percentages of published documents. Nowadays, authors have an interest in publishing in journals with the highest possible impact factor, and, therefore, it is reasonable to suppose that first-quartile journals accept more documents than fourth-quartile journals. This supposition bear out bear out previous work, like Cabanac (2012), stating that the range of papers accepted per journal is wider for the first-quartile than for the other quartiles.

According to the distribution of document published in different quartiles during the analyzed period, we observed that first-quartile had the highest percentage (30.490 %), followed by second-quartile (27.460 %), third-quartile (27.335 %) and fourth-quartile (14.715 %). The evolution of documents published in different quartiles is analyzed in Fig. 10. We noted that the highest percentage of first-quartile documents were achieved in 2009, whereas the highest percentage of fourth-quartile documents were achieved in 2002. Analyzing Fig. 10, we also observed that the percentage of documents published by first- and second-quartile journals have gradually increased, whereas the percentage of documents published by third- and fourth-quartile journals have decreased.

Table 5 Percentages of documents published in each quartile by different numbers of authors

| Authors	First-quartile (%)	Second-quartile (%)	Third-quartile (%)	Fourth-quartile (%)
1	28.333	24.167	31.667	15.833
2	27.086	24.305	33.821	14.788
3	30.037	28.189	27.726	14.048
4	32.392	29.032	23.387	15.189
5	31.268	28.024	23.304	17.404
>5	36.481	29.185	22.747	11.587
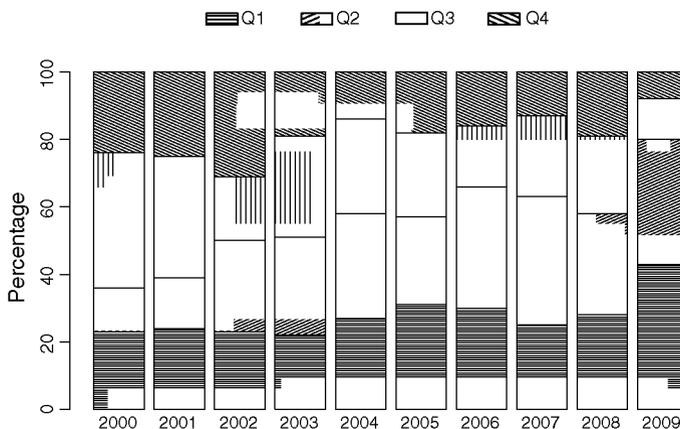
Fig. 10 Evolution of percentages of documents published in quartile journals

The evolution of documents published in different quartiles according to different numbers of authors is analyzed in Fig. 11. We found that there were small differences in publication behavior by author set cardinality for documents across different quartiles. In general, we found that documents published by one and two authors have undergone a percentage decrease, leading to a drop in the collaborative rate value throughout the analyzed period. We also found that the percentage of documents published by four authors has increased, whereas documents written by three, five and more than five authors have undergone fluctuations and small decreases with respect to different quartiles. As expected, documents published by three or four authors had the highest values in each quartile. There has been a noteworthy increment of documents published by four authors in last few years, rising to 47.252 % of documents in third-quartile journals (see Fig. 11c) and 55.814 % of documents in fourth-quartile journals (see Fig. 11d) in 2009. On the other hand, the main differences were associated with documents written by three authors. In this case, we found that the percentages of these documents fluctuated in the first- and second-quartile journals, whereas they clearly decreased in third- and fourth-quartile journals.

Figure 12 analyzes the collaborative level with respect to different quartiles. It shows that fourth-quartile journals had the highest value for number of authors per document (4.519 authors) in 2009, followed by journals belonging to the first (3.869 authors), third (3.587 authors) and second (3.569 authors) quartiles. We found that all collaborative level values have undergone an increase in the last few years. Especially noteworthy was the sizeable increment of fourth-quartile journals since 2006. According to these results, we believe that fourth-quartile journals have recently accepted documents with many authors (see Fig. 11d) in order to improve their number of citations. A possible interpretation is that these journals publish documents with many co-authors in order to improve their quartile through increased dissemination by co-authors including self-citations.

Table 6 presents the average number of citations per document, the average age per document, and the average number of citations per document and year. These values and their standard deviations were calculated for documents belonging to each quartile. As expected, we found that documents published by first-quartile and second-quartile journals usually had a higher average number of citations per document and year than documents published by third-quartile and fourth-quartile journals.
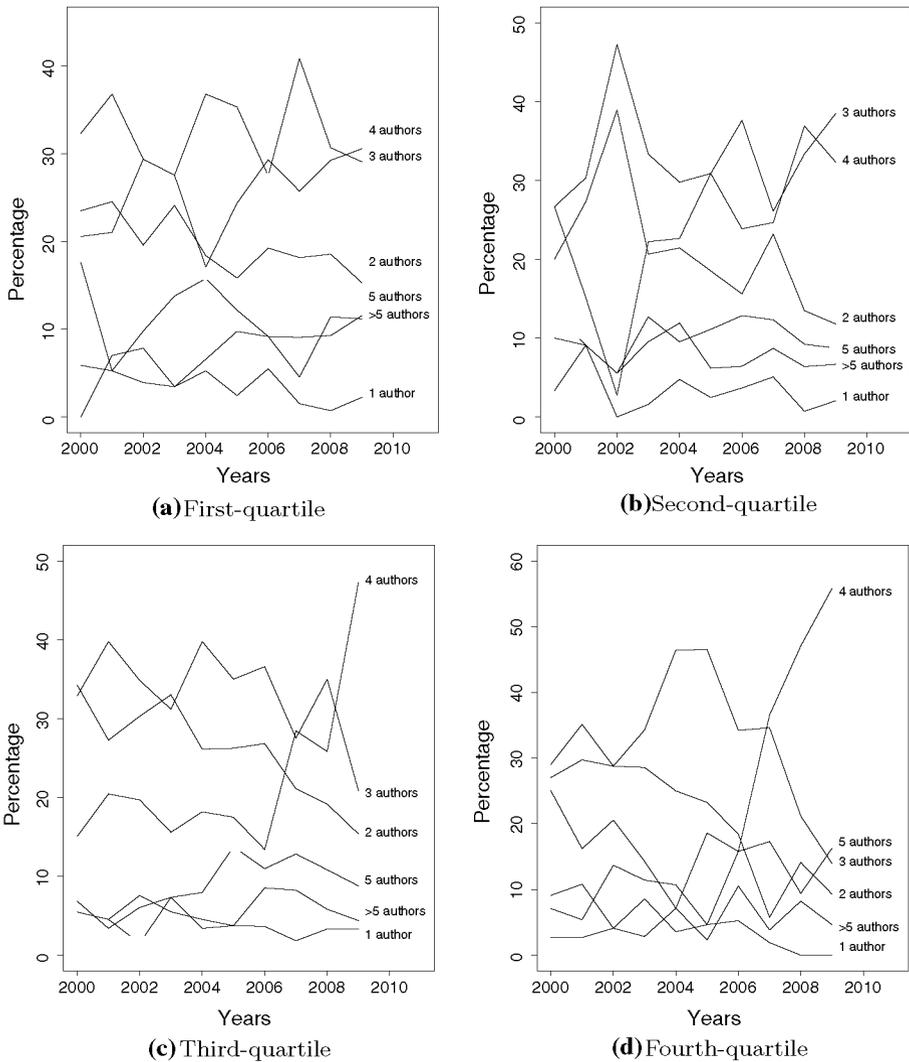
Fig. 11 Evolution of percentages of documents published in quartile journals by number of authors

Analyzing the number of authors, we noted that single-authored documents always had the lowest average value of citations per document and year when they were published by first-, second- and third-quartile journals. In contrast, these documents had the highest average value (1.100 ± 1.734 citations) when they were published in fourth-quartile journals. We performed a Kruskal-Wallis test in order to compare subsets of different authors across documents published in different quartiles. Results show that there were no significant differences across documents published by first- and third-quartile journals. In contrast, it found significant differences across documents published by second- and fourth-quartile journals: the average number of citations per document and year of documents published by five authors in second-quartile journals (1.118 ± 1.466) was significant different (higher) to documents published by one author (0.460 ± 0.816), two authors
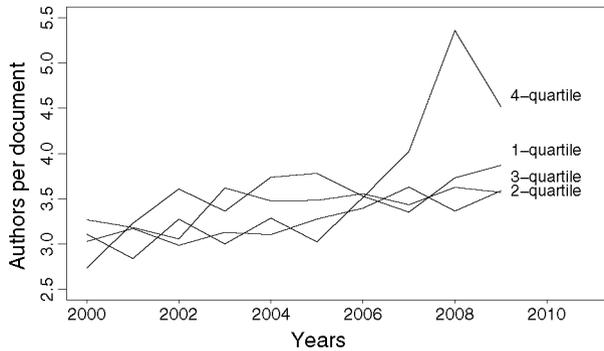
Fig. 12 Evolution of the average number of authors per document (CL) by journal impact factor quartiles

Table 6 Mean ± standard deviation for citation measures of documents published by numbers of authors according to impact factor

	First-quartile	Second-quartile	Third-quartile	Fourth-quartile
1-author				
Citation count	6.176 ± 11.862	2.069 ± 3.240	4.237 ± 7.713	8.789 ± 15.183
Publication age	5.559 ± 3.193	4.862 ± 2.900	5.947 ± 2.837	7.263 ± 2.207
Citations ratio	0.815 ± 1.255	0.460 ± 0.816[†]	0.574 ± 0.938	**1.100 ± 1.734**
2-authors				
Citation count	8.811 ± 21.678	5.000 ± 12.819	5.723 ± 14.828	3.442 ± 5.775
Publication age	4.227 ± 2.837	4.530 ± 2.625	5.792 ± 2.763	6.253 ± 2.737
Citations ratio	**1.474 ± 2.717**	0.836 ± 1.634[†]	0.794 ± 1.578	0.454 ± 0.715
3-authors				
Citation count	6.708 ± 13.177	3.859 ± 7.928	5.683 ± 11.643	2.736 ± 3.679
Publication age	3.969 ± 2.707	3.833 ± 2.574	5.467 ± 2.776	5.521 ± 2.660
Citations ratio	1.359 ± 2.167	0.767 ± 1.219[†]	0.860 ± 1.380	0.452 ± 0.589
4-authors				
Citation count	6.178 ± 10.491	4.495 ± 8.462	4.845 ± 9.487	2.163 ± 4.451
Publication age	3.859 ± 2.701	3.662 ± 2.413	4.891 ± 2.945	4.750 ± 3.052
Citations ratio	1.234 ± 1.654	1.009 ± 1.641	0.744 ± 1.152	0.374 ± 0.618
5-authors				
Citation count	4.991 ± 7.321	5.684 ± 8.714	5.000 ± 7.402	2.389 ± 4.866
Publication age	3.755 ± 2.570	4.021 ± 2.497	4.418 ± 2.520	4.778 ± 2.879
Citations ratio	1.037 ± 1.265	**1.118 ± 1.466**	0.936 ± 1.209	0.390 ± 0.632
>5-authors				
Citation count	5.176 ± 8.181	3.853 ± 4.643	5.962 ± 16.233	1.042 ± 2.053
Publication age	3.200 ± 2.429	4.044 ± 2.464	5.113 ± 2.819	4.375 ± 2.732
Citations ratio	1.153 ± 1.465	1.029 ± 1.959	**0.975 ± 2.135**	0.238 ± 0.488[†]

[†] Represents those results that are statistically different in citations ratio with respect to the benchmark subset (highlighted in boldface)

(0.836 ± 1.634) and three authors (0.767 ± 1.219). Likewise, the average number of citations per document and year of documents published by one author in fourth-quartile journals (1.100 ± 1.734) was significant different (higher) to documents published by more than five authors (0.238 ± 0.488). According to these results, we found no pattern to explain the relationship between impact factors, utility and authors.

Discussion and conclusions

Let us emphasize that our analysis is carried out for one nation only: Spain. It is also limited to the research production included in the Web of Science, so some national conferences and journals (which are a few) in Spanish are not taken into account. We analyzed a small percentage of the worldwide output, therefore, the results may not be generally applicable. Further research is required in order to assess the above questions.

   We know that the number of citations could vary depending on each database (Web of Knowledge, Scopus and Google Scholar, etc.) (Bar-Ilan 2008). The Web of Knowledge, which is the database consulted in this study, stores the most relevant scientific literature produced and published worldwide in different areas of knowledge and disciplines (Garfield 2003). The prestige associated with the Web of Knowledge is the stringent selection criteria applied to journals and conferences. These rigorous selection processes are supported by bibliometric laws, which show that the best science is found in small clusters (Garfield 2000). Despite Web of Knowledge flaws in the computer science field (Wainer et al. 2011), this platform has a specialized conference proceedings database (Conference Proceedings Citation Index) which stores 400,000 publications annually from more than 15,000 different computer science conferences. Additionally, the Web of Knowledge includes the most important databases specialized in journal articles (Science Citation Index and Journal Citation Reports) covering around 450 computer science journals.

   This paper has studied five relationships. The first analyzes how productivity and utility vary according by number of authors. According to productivity, our initial hypothesis was that the average computer science document was written by three or four authors. The research findings confirm that our hypothesis was correct. Results also show that the collaborative level has increased over time as expected. This was caused by both the percentage decrease of published documents written by one and two authors, and the percentage increase of documents written by three or more authors. On the other hand, we believed that a higher number of authors would lead to a higher number of citations. In contrast, results show that documents published by one or two authors have higher values of citations per document and year than documents published by three or more authors. In fact, statistical test results show that there are significant differences between the 2-author subset and subsets with more authors. We did not find a positive association between author set cardinality and the citation impact.

   The second relationship analyzes how productivity and utility vary across different types of collaboration according to different number of authors. Due to the problems concerning geographic distance and communication across organizations, we believed that most documents were written via institutional collaboration. Results show that the initial hypothesis was correct, since the 73.554 % of total publications were published by authors belonging to the same institution. On the other hand, we found that publication behavior was similar across different types of collaboration. International, national and institutional collaborations are usually written by three or four authors. Regarding collaborative level,

we thought that international documents would have more authors than national and institutional documents. Results show that the initial hypothesis was again correct. A possible explanation of this fact is that a large number of international and national collaborations spring from projects that require the participation of many researchers at different institutions, whereas most institutional collaborations usually involve authors from the same research group, that is, involve fewer authors. Finally, we believed that international collaborations would have more citations than national and institutional documents. Unlike Bartneck and Hu (2010) who were unable to find a general beneficial effect of collaboration of any type (international, national or institutional) on the quality of the papers measured by their citation counts, we found that international collaborations always have the highest average number of citations per document and year for different numbers of authors. We also believed that a greater number of authors would lead to a higher number of citations with a particular type of collaboration. However, statistical test results show that document utility is better if it is published by few authors at the institutional level, whereas the number of authors does not affect citation counts at national and international level.

The third relationship investigates how productivity and utility vary across different document types according to different number of authors. We believed that the publication rate associated with journal articles and conference papers would be different. Results corroborated this hypothesis, showing that 32.262 % of publications belong to journal articles, whereas 67.738 % of publications belong to conference papers. Analyzing the collaborative level, our initial hypothesis was correct. We found that nowadays there are no important differences in publication behavior between journals and conferences as believed, but we also found that the collaboration level in conference papers was higher than journal articles in earlier years. We noted that there is a general decrease of documents published by one and two authors in journal and conference publications. According to the number of authors, we noted that single-authored documents are usually published in journals, whereas multi-authored documents are usually published in conferences. We also believed that citation counts would be different between journals and conferences. Results show that journal articles have more citations per document and year than conference papers. Finally, unlike our initial hypothesis, statistical test results do not assure that a greater number of authors leads to more citations.

The fourth relationship investigates how productivity and utility among the seven computer science subdisciplines vary by number of authors. We believed that the publication rate associated with each subdiscipline would be different. Results corroborated this hypothesis, showing that 40.123 % of publications belong to theory and methods, whereas only 1.613 % of publications belong to cybernetics. Regarding the percentages of documents published by different numbers of authors, we find that there are no important differences in publication behavior across subdisciplines as believed. We find that there is a general decrease of documents published by one and two authors in all subdisciplines. Also, we find that three and four authors write the highest percentage of documents in all subdisciplines, whereas documents written by one author account for the the lowest percentage. Analyzing the collaborative level, our initial hypothesis was correct. We also believed that citation counts would be different across subdisciplines. Results show that documents related to artificial intelligence, cybernetics and interdisciplinary applications usually have the highest value of citations per document and year with a set number of authors. Finally, unlike our initial hypothesis, statistical test results do not assure that a greater number of authors leads to higher number of citations within a specific a subdiscipline.

The last relationship analyzes how productivity and utility in different journal impact factor quartiles vary by number of authors. We believed that first-quartile journals would have the lowest percentage of publications due to their low acceptance rate. Contrariwise, results show that first- and second-quartile journals publish more documents than third- and fourth-quartile journals. This is reasonable bearing in mind that authors have an interest in publishing in journals with the highest possible impact factor nowadays. Regarding the number of authors, we supposed that there would be no clear differences across quartiles. In contrast, we find an important increment of number of authors per documents in fourth-quartile journals since 2006. As expected, results show that citation counts are obviously different across quartiles. Finally, we also believed that a greater number of authors would lead to a higher number of citations for a set quartile. However, statistical test results found no pattern to explain the relationship between impact factor quartiles, number of citations and number of authors.

In the future, our target will be to analyze other aspects related to collaboration at author level (number of different co-authors, productivity of co-authors, utility of co-authors, proximity among co-authors, etc). We are interested in analyzing the characteristics of a specific author's co-authors. Also, we will analyze whether researchers with the best research performance are also the investigators that collaborate more at the international level, and whether the citation counts of papers that have been written by authors with a low number of citations improve through collaboration.

References

Abramo, G., D'Angelo, C., & Solazzi, M. (2011a). Are researchers that collaborate more at the international level top performers? An investigation on the Italian university system. *Journal of Informetrics 5*(1), 204–213.

Abramo, G., D'Angelo, C., & Solazzi, M. (2011b). The relationship between scientists' research performance and the degree of internationalization of their research. *Scientometrics 86*(3), 629–643.

Archibugi, D., & Coco, A. (2004). International partnerships for knowledge in business and academia: a comparison between Europe and the USA. *Technovation 24*(7), 517–528.

Bammer, G. (2008). Enhancing research collaborations: three key management challenges. *Research Policy 37*(5), 875–887.

Bar-Ilan, J. (2008). Which h-index? A comparison of WoS, Scopus and Google Scholar. *Scientometrics 74*(1), 257–271.

Bartneck, C., & Hu, J. (2010). The fruits of collaboration in a multidisciplinary field. *Scientometrics 85*(1), 41–52.

Beaver, D. (2001). Reflections on scientific collaboration (and its study): past, present and future. *Scientometrics 52*(3), 365–377.

Beaver, D., & Rosen, R. (1979). Studies in scientific collaboration. Part I. The professional origins of scientific co-authorship. *Scientometrics 1*(1), 65–84.

Bornmann, L., & Daniel, H. (2006). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation 64*(1), 45–80.

Cabanac, G. (2012). Shaping the landscape of research in information systems from the perspective of editorial boards: a scientometric study of 77 leading journals. *Journal of the American Society for Information Science and Technology 63*(5), 977–996.

Cronin, B. (1981). The need for a theory of citing. *Journal of Documentation 37*(1), 16–24.

Cullen, P., Norris, R., Resh, V., Reynoldson, T., Rosenberg, D., & Barbour, M. (1999). Collaboration in scientific research: a critical need for freshwater ecology. *Freshwater Biology 42*(1), 131–142.

Fortnow L. (2009) Time for computer science to grow up. *Communications of the ACM 52*(8), 33–35.

Franceschet, M. (2010). The role of conference publications in computer science: a bibliometric view. *Communications of the ACM 53*(12), 129–132.

Franceschet M. (2011) Collaboration in computer science: a network science approach. *Journal of the American Society for Information Science and Technology 62*(10), 1992–2012.

Franceschet, M., & Costantini, A. (2010). The effect of scholar collaboration on impact and quality of academic papers. *Journal of Informetrics 4*(4), 540–553.

Garfield, E. (2000). Use of journal citation reports and journal performance indicators in measuring short and long term journal impact. *Croatian Medical Journal 41*(4), 368–374.

Garfield, E. (2003). The meaning of the Impact Factor. *International Journal of Clinical and Health Psychology 3*(2), 363–369.

Gazni, A., Sugimoto, C., & Didegah, F. (2012). Mapping world scientific collaboration: authors, institutions, and countries. *Journal of the American Society for Information Science and Technology 63*(2), 323–335.

Glanzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics 51*(1), 69–115.

Hauptman, R. (2005). How to be a successful scholar: publish efficiently. *Journal of Scholarly Publishing 36*(2), 115–119.

Ibañez, A., Bielza, C., & Larrañaga, P. (2012). Analysis of scientific activity in Spanish public universities in the area of computer science. *Revista Española de Documentacion Cientifica* (accepted).

Katz, J., & Martin, B. (1997). What is research collaboration? *Research Policy 26*(1), 1–18.

Kruskal, W., & Wallis, W. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association 47*(260), 583–621.

Laine, C., & Mulrow, C. (2005) Exorcising ghosts and unwelcome guests. *Annals of Internal Medicine 143*(8), 611–612.

Lancho-Barrantes, B., Guerrero Bote, V.P., Chinchilla-Rodriguez, Z., & de Moya Anegon, F. (2012). Citation flows in the zones of influence of scientific collaborations. *Journal of the American Society for Information Science and Technology 63*(3), 481–489.

Landry, R., & Amara, N. (1998). The impact of transaction costs on the institutional structuration of collaborative academic research. *Research Policy 27*(9), 901–913.

Levitt, J., & Thelwall, M. (2009). Citation levels and collaboration within library and information science. *Journal of the American Society for Information Science and Technology 60*(3), 434–442.

Liao, C. (2011). How to improve research quality? Examining the impacts of collaboration intensity and member diversity in collaboration networks. *Scientometrics 86*(3), 747–761.

Mann, H., & Whitney, D. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics 18*(1), 50–60.

Olson, G., & Olson, J. (2000). Distance matters. *Human Computer Interaction 15*(2), 139–179.

Persson, O., Glanzel, W., & Danell, R. (2004). Inflationary bibliometric values: the role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics 60*(3), 421–432.

Ponomariov, B., & Boardman, P. (2010). Influencing scientists' collaboration and productivity patterns through new institutions: University research centers and scientific and technical human capital. *Research Policy 39*(5), 613–624.

Presser, S. (1980). Collaboration and the quality of research. *Social Studies of Science 10*(1), 95–101.

Ruiz-Perez, R., Delgado-Lopez-Cozar, E., & Jimenez-Contreras, E. (2002). Spanish personal name variations in national and international biomedical databases: implications for information retrieval and bibliometric studies. *Journal of the Medical Library Association 90*(4), 411–430.

Smolinsky, L., & Lercher, A. (2012). Citation rates in mathematics: a study of variation by subdiscipline. *Scientometrics 91*(3), 911–924.

Solomon, J. (2009). Programmers, professors, and parasites: credit and co-authorship in computer science. *Science and Engineering Ethics 14*(5), 476–489.

Sooryamoorthy, R. (2009). Do types of collaboration change citation? Collaboration and citation patterns of South African science publications. *Scientometrics 81*(1), 177–193.

VanRaan, A. (1998). The influence of international collaboration on the impact of research result. *Scientometrics 42*(3), 423–428.

Wainer, J., Billa, C., & Goldenstein, S. (2011). Invisible work in standard bibliometric evaluation of computer science. *Communications of the ACM 54*(5), 141–146.

Weingart, P. (2005). Impact of bibliometrics upon the science system: inadvertent consequences? *Scientometrics 62*(1), 117–131.

Zetterstrom, R. (2004). The number of authors of scientific publications. *Acta Paediatrica 93*(5), 581–582.