Measuring the Impact of the International Relationships of the Andalusian Universities Using Dimensions Database

P. García-Sánchez[(✉)] and M. J. Cobo

Department of Computer Science and Engineering, University of Cádiz, Cádiz, Spain
{pablo.garciasanchez,manueljesus.cobo}@uca.es

Abstract. Researchers usually have been inclined to publish papers with close collaborators: same University, region or even country. However, thanks to the advancements in communication technologies, members of international research networks can cooperate almost seamlessly. These networks usually tend to publish works with more impact than the local counterparts. In this paper, we try to demonstrate if this assumption is also valid in the region of Andalusia (Spain). The Dimensions.ai database is used to obtain the articles where at least one author is from an Andalusian University. The publication list is divided into 4 geographical areas: local (only one affiliation), regional (only Andalusian affiliations), national (only Spanish affiliations) and International (any affiliation). Results show that the average number of citations per paper increases as the author collaboration networks increases geographically.

Keywords: Bibliometric analysis · International collaboration
Andalusian universities · Dimensions.ai

1 Introduction

Science is usually developed in teams that could be considered domestic, if the researchers are all from the same country, or international, if researchers belong to different countries. According to [1], the fourth age of research is driven by international collaborations, which could be motivated by two kinds of factors [2] related to the diffusion of scientific capacity, or related to the interconnectedness of researchers. That is, in the current global knowledge society [3], researchers tend to collaborate with colleagues from other countries in order to advance in their own fields.

Furthermore, papers developed within international teams, used to be more cited than paper developed within domestic teams [2,4]. Indeed, recently has been demonstrated that researchers to develop their career in different countries tend to be more cited that those who still in the same country in all their academic life [5].

So, the main aim of this contribution is to determine if there an increase in the number of citations when researchers from different geographical areas

collaborate. To answer this question, we develop a bibliometric analysis [6–8] we have focused on the 9 Universities from the region of Andalusia (Spain). We compare the number of citations in the next scenarios: when the publications are only signed by members of the same Andalusian university, when they are only signed by members of Andalusian Universities, when they are signed by members of Andalusian Universities and other affiliations from Spain, and finally when they are signed by at least one member from an Andalusian University. The number of publications, citations and citations per publication is compared.

The Dimensions database has been used to obtain the publications and number of citations. We have chosen this database because it is freely available for academic purposes, it includes a large corpus (i.e., 89m publications and 4 billion references), and it has a powerful API that allows advanced analytic by means of their own DSL (Domain Specific Language) query language, and using a programming language such as Python.

The rest of the paper is structured as follows: in Sect. 2 the methodology to obtain the dataset of publications is explained. Then, a discussion of the results is presented in Sect. 3. Finally, conclusions and future work are presented.

2   Methodology

This section describes the steps followed to obtain the dataset of articles published by researchers from Andalusian Universities. The data is sourced from Dimensions, an inter-linked research information system provided by Digital Science (https://www.dimensions.ai). This database has been chosen not only because its large amount of data available, including the number of citations by publication, but also because it offers the possibility to use an API to perform queries using a DSL. This SQL-like petitions can be performed from any programming language and used to obtain a large batch of specific results in JSON format, thus facilitating their processing and analysis. Python language has been used to call the Dimensions API and to plot and analyse the results.

The data we are interested in are the publications signed by at least one member from one Andalusian University. As we want to compare the citation number when different regional and international collaborations appear, we divided the dataset as follows, from local to international co-authoring:

- P_{One}: Papers from Andalusia (only one affiliation). This dataset includes the papers where all the authors belong to the same Andalusian University.
- P_{And}: Papers from Andalusia (only with Andalusian Universities collaborators). This dataset includes the papers where all the authors belong to Andalusian Universities.
- P_{Spa}: Papers from Andalusia (only with Spanish Entities collaborators). This dataset includes the papers where all authors belong to a Spanish institution, and at least one is from Andalusia.
- P_{All}: All papers from Andalusia universities. All papers where at least one author is from a University from Andalusia.

Every dataset is included in the one above, therefore $P_{One} \subseteq P_{And} \subseteq P_{Spa} \subseteq P_{All}$.

The rest of parameters for the queries are:

– The selected universities to perform the analysis are the 9 public universities of Andalusia. To perform the queries, their associate Global Research Identifier Database (GRID), available from https://grid.ac/, has been used[1].
– Date range: from 2010 to 2015.
– Only publications of type "article" are used.
– The queries were performed on 25th July 2018.

To obtain all the papers from Andalusia Universities (P_{All}), the query used is:

```
search publications
 where (year in [2010:2015]
 and research_orgs.id in
  ["grid.4489.1","grid.7759.c","grid.9224.d","grid.15449.3d","grid.18803.32",
   "grid.411901.c","grid.10215.37","grid.21507.31","grid.28020.38"]
 and type="article"
return publications[id+year+times_cited+research_orgs+FOR_first+funders]
sort by id limit 1000 skip 0
```

The query to obtain the papers of P_{One} directly from dimensions, the modifier `and count(research_orgs) = 1` can be used:

```
search publications
 where (year in [2010:2015]
 and research_orgs.id in
  ["grid.4489.1","grid.7759.c","grid.9224.d","grid.15449.3d","grid.18803.32",
   "grid.411901.c","grid.10215.37","grid.21507.31","grid.28020.38"]
 and count(research_orgs) = 1)
 and type="article"
return publications[id+year+times_cited+research_orgs+FOR_first+funders]
sort by id limit 1000 skip 0
```

To select all the papers from Andalucia Universities with only collaborators from Spain, the modifier used is `count(research_org_countries) = 1`.

```
search publications
 where (year in [2010:2015]
 and research_orgs.id in
  ["grid.4489.1","grid.7759.c","grid.9224.d","grid.15449.3d","grid.18803.32",
   "grid.411901.c","grid.10215.37","grid.21507.31","grid.28020.38"]
 and count(research_org_countries) = 1
 and type="article"
return publications[id+year+times_cited+research_orgs+FOR_first+funders]
sort by id limit 1000 skip 0
```

---

[1] The list of Universities and their and associated id are: University of Cádiz (grid.7759.c), University of Sevilla (grid.9224.d), Pablo de Olavide University (grid.15449.3d), University of Huelva (grid.18803.32), University of Córdoba (grid.411901.c), University of Málaga (grid.10215.37), University of Jaén (grid.21507.31) and University of Almería (grid.28020.38).

Dimensions does not offer the possibility to filter directly using "only from" a specific attribute (for example, only from the list of universities). That is the reason we have filtered the P_{All} dataset to obtain P_{And} iterating for every publication and removing those that have authors whose affiliation is not in the Andalusian universities list.

3   Results

Obtained results and corpus sizes are summarized in Table 1. As it can be seen, a great number of articles (39.81%) are only signed by researchers of the same affiliation (P_{One}). On the other side, Andalusian researchers tend to collaborate with researchers from affiliations outside Andalusia: 1426 publications with other Andalusian Universities, and 5536 with other Spanish affiliations (without counting Andalusian Universities, that is $P_{Spa} - P_{And}$). But it is remarkable that 13944 publications ($P_{All} - P_{Spa}$) are signed by Andalusian Universities and foreign affiliations (40.14% of the total), more than with regional and national collaborators. Results in Table also show that the average number of citations increases when including publications from other affiliations, clearly being the P_{All} dataset the one with the larger value (15.675). This can be explained because large research projects involving different countries are more ambitious than regional ones.

Table 1. Summary of citations and publications per corpus. All publications are included in the dataset below: $P_{One} \subseteq P_{And} \subseteq P_{Spa} \subseteq P_{All}$

Corpus	#Publications	#Citations	Average citations per publication
P_{One}	13,832	162,211	11.727
P_{And}	15,258	185,525	12.159
P_{Spa}	20,794	265,997	12.792
P_{All}	34,738	544,559	15.676

Plotting the citation histogram of all publications also show clear differences between the datasets. Figure 1 shows that the publication citations follow a long-tail scheme, where the majority of the publications are not cited, or below of 200 citations, while a few highest cited papers are cited up to 594 times. When increasing the geographical ambit, higher cited papers appear, for example for P_{And} and P_{Spa} (Figs. 2 and 3), the highest cited paper has 1089 citations, clearly being a paper from P_{And}. However, the average citation per paper is still greater if we take into account the Spain geographical ambit. Although another highly cited paper appears in Fig. 4 with 839 citations, the differences between both datasets are not so clear. It is when plotting all the papers with Andalusian authors and the rest of the world P_{All} (Fig. 4), where the highest cited papers appears (4231, 2919, 2309 and 1515 respectively), but also, the group at the beginning of the x-axis moves to higher amount of citations.
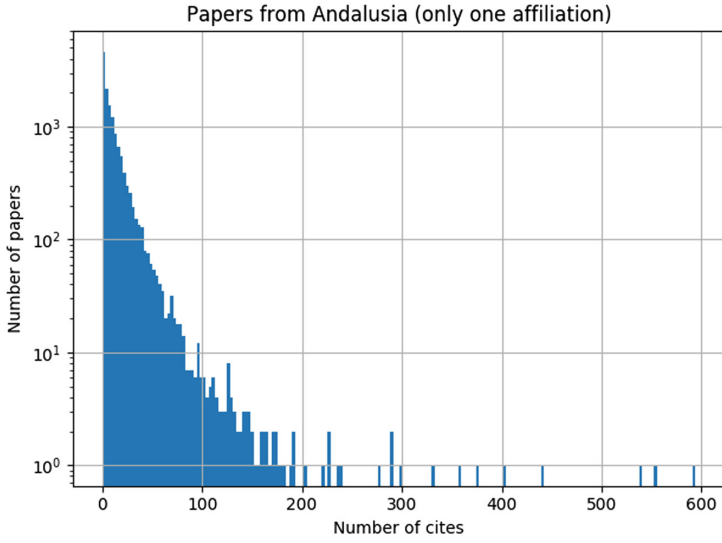
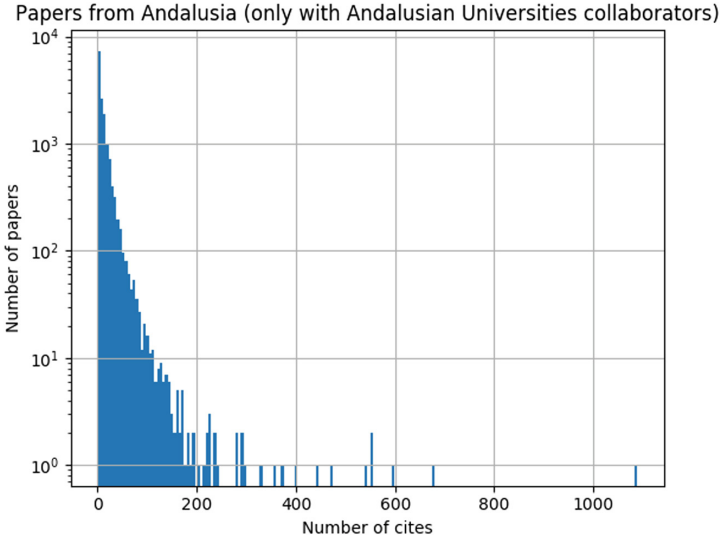**Fig. 1.** Citation histogram for P_{One}. Y-axis uses a logarithmic scale.

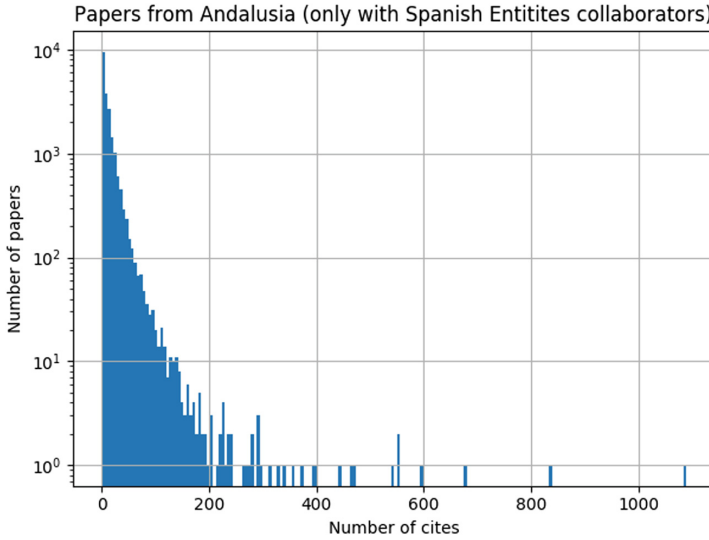**Fig. 2.** Citation histogram for P_{And}. Y-axis uses a logarithmic scale.

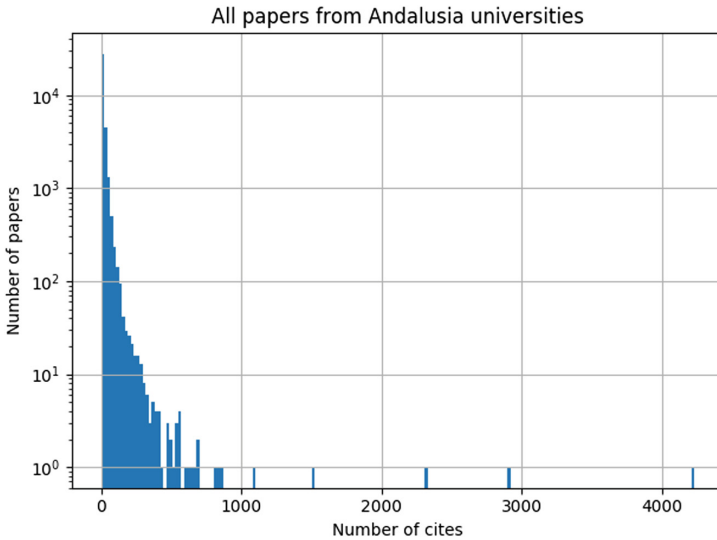**Fig. 3.** Citation histogram for P_{Spa}. Y-axis uses a logarithmic scale.

Fig. 4. Citation histogram for P_{All}. Y-axis uses a logarithmic scale.

4   Conclusions

In this paper, the number of citations of articles from Andalusian Universities is analysed taking into account the geographical collaboration network of the authors. The Dimensions.ai database has been used to obtain the articles from

Andalusian researchers, divided into 4 different geographical areas: only one affiliation, only Andalusian affiliations, only Spanish affiliations and all publications.

Results show that Andalusian publications are clearly divided into two groups: articles signed by researchers within the same affiliation (39.8%) and signed with researchers from foreign countries (40.14%). The average number of citations per paper also increases when the collaboration network geographically increases, meaning that publications with international collaborations obtain more citations than the ones with only one affiliation.

Future studies will include more complete information by separating the presented datasets into disjoint sets, or limiting by specific University. Another kind of analysis, such as collaboration graphs between countries or universities, may give more insight to determine the members of each network or the quality of the research. Furthermore, Dimensions API also allows to obtain the funding agencies of each publication, so a study to compare the impact of projects funded by different countries can be performed. Patents and clinical trials are also available in Dimensions, so a comparison of the different types of publications may also be relevant to the issue.

References

1. Adams, J.: The fourth age of research. Nature **497**(7451), 557–560 (2013)
2. Wagner, C.S., Leydesdorff, L.: Network structure, self-organization, and the growth of international collaboration in science. Res. Policy **34**(10), 1608–1618 (2005)
3. Moed, H.F., Aisati, M., Plume, A.: Studying scientific migration in Scopus. Scientometrics **94**(3), 929–942 (2013)
4. Persson, O., Glänzel, W., Danell, R.: Inflationary bibliometric values: the role of scientific collaboration and the need for relative indicators in evaluative studies. Scientometrics **60**(3), 421–432 (2004)
5. Sugimoto, C.R., Robinson-Garcia, N., Murray, D.S., Yegros-Yegros, A., Costas, R., Larivière, V.: Scientists have most impact when they're free to move. Nature **550**(7674), 29–31 (2017)
6. Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E., Herrera, F.: Science mapping software tools: review, analysis, and cooperative study among tools. J. Am. Soc. Inf. Sci. Technol. **62**(7), 1382–1402 (2011)
7. Fortunato, S., et al.: Science of science. Science **359**(6379) (2018)
8. Gutiérrez-Salcedo, M., Martínez, M.Á., Moral-Munoz, J.A., Herrera-Viedma, E., Cobo, M.J.: Some bibliometric procedures for analyzing and evaluating research fields. Appl. Intell. **48**(5), 1275–1287 (2018)