Scientometrics (2020) 123:635–653
https://doi.org/10.1007/s11192-020-03399-4

Science, research, and innovation infospheres in Google results of the Ibero-American countries

Simone Belli[1] · **Carlos Gonzalo-Penela**[2]

Received: 6 March 2019 / Published online: 7 March 2020

Abstract
The growing digitalization of scientific research practices is reflected in the content that academic and governmental institutions put on their websites, many of which are not optimized so that their contents reach visibility in search results of Google. Through the mapping of search engine results, this article analyzes the visibility of Ibero-American governmental, educational and research institutions in the results of Google in relation to a group of keywords related to the areas of Science, Research and Innovation. By analyzing the results of these pages in the search results in a specific period we can determine that, although few exceptions, the algorithms used by Google increase the visibility of educational and research institutions in Ibero-America (IA) along with those of each country in function of the national search option offered by the search engine. The indicators obtained both for web presence and web visibility indicate that pages that appear more frequently in the first positions in IA countries are not owned by national institutions, but from other countries. Moreover, we have observed that governmental and educational institutions are most visible than research institutions. While previously social networks are not so far popular for this type of institutions, they are recently gaining positions. However, this study is exploratory and a longitudinal research would eliminate fluctuations of web data.

Keywords Web mapping · Universities · Research centers · Visibility · Google · Adwords · SERPs

Mathematics Subject Classification 91C05 Game theory · Economics · Social and behavioral sciences · Social and behavioral sciences: general topics · Measurement theory

JEL Classification C80 data collection and data estimation methodology · Computer programs · General

✉ Simone Belli
sbelli@ucm.es

Carlos Gonzalo-Penela
carlos.gonzalo@upf.edu

1 Department of Social Anthropology and Social Psychology, Faculty of Political Science and Sociology, Complutense University of Madrid, Madrid, Spain

2 Pompeu Fabra University, Barcelona, Spain

Introduction

In this study, we map the main Webpages related to Science, Research, and Innovation (SRI) in Ibero-American (IA) countries. We have launched an informational search in Google of 20 keywords in Spanish based on SRI focusing mainly on universities and research institutes. We analyze the associated data visibility and the kind of websites listed by Google for each country. Consequently, this study will seek an answer to our research questions (RQ), directed to evaluate the relevance of the data obtained as an indicator of digital practices displayed by institutions according to the Search Engine Optimization (SEO) visibility criteria:

RQ1: What is the composition of Google results sites in IA when we conduct searches related to SRI, and, does Google apply the same proportions in all IA countries?
RQ2: Are there significant differences between governmental, educational (universities) or research institutions in the web?
RQ3: What kind of pages appear more frequently in the first positions? (University and research group's homepages, SRI public or private institutions, scientific papers, etc.)
RQ4: Are the most visible websites for each IA country national or foreign organizations?
RQ5: What is the relative visibility impact of news websites and social media websites on SRI-related searches?
RQ6: What role play university websites in this emerging scenario?

This paper attempts to gain insight into the stability of results for webometric studies and deals specifically with trends of SRI-IA websites. Measuring the impact and the visibility that these keywords have on the web is the first step to map the presence of universities based on search queries related to SRI.

We have focused our analysis on IA countries because these represent a specific and emerging scientific context to study and an unknown territory until now. Shared language and similar culture have favored mobility of students and researchers among IA countries. International scientific collaboration rates of IA countries observing scientific co-publications show significant networking practices in the last decade (Belli and Baltà 2019; Minniti et al. 2018) that should be reflected on their SRI Websites. Despite this in most cases, the results shown by Google improve access to the contents of each country acting as a detriment to other content published by institutions in other neighboring countries. This is because of two factors, one being the search engine criteria when it comes to showing results, and the other the behavior of national users in each country. To prove it, we apply descriptive statistics analysis to Google Search Engine Result Pages, from now on abbreviated using the acronym SERPs. We have discovered what elements are composed, and identified which role SRI-IA institutions play in the results of the most used search engine worldwide.

Previous related research

Nowadays, researchers in different disciplines have a wide variety of options to share their results and data. Research organizations and scholars are affected by the digital transformation in research. This new context is defined as e-science and cyberscience (Beaulieu 2001; Hine 2002; Berman 2004; Nentwich 2003). The first term stresses the practice of research, while the second one underlines the infrastructural condition for that practice. However, both concepts refer to a shared view of digital research as a new way of doing research.

Besides, digital transformation has changed the way academic community communicates. Digital communication and interaction have distorted geography by shrinking distances and removing access barriers. Nowadays, scholars need to be experts in communicating their research, increasing visibility on the Internet is a key factor to success. Visibility is defined as the extent to which a user is likely to come across a reference to a Website (Drèze and Zufryden 2004). In many cases, the connectivity of a researcher in a scientific network is more important than the quality of research. For Callon and Law (2004), circulation has become more important than fixed positions for this reason. e-Science has increased collaboration within institutions and disciplines creating a central infrastructure (Ackland et al. 2007).

In this scenario, academic web networks are crucial to share research results between different researchers and institutions. Thelwall (2000) identifies how academic domains tend to be better linked than commercial domains, and how university and research groups webpages are essential. Nevertheless, many times the high link counts are not associated with the quality of scholar content (Thelwall 2002). Thelwall and Harries (2004) have demonstrated that universities that conduct more research attract significantly more links. (Arslan and Seker 2014) have examined the tools of reputation of academic websites, and the relation between online reputation and effectiveness of the universities.

There is a large number of research studies (Dominic and Jati 2010; Kaur et al. 2016; Olaleye et al. 2018) that to develop their rankings use technological criteria, size, and performance of the analyzed sites. These criteria coincide with what we could consider onpage positioning factors of the SEO specialty area (Codina et al. 2017). Among these analyses that use site's own criteria, (Orduna-Malea et al. 2015) have focused their research on Google Scholar, which is a pure academic search engine depending directly on World-Wide Web coverage, measuring its size. This type of web consists on trusted websites that host academic materials such as universities, repositories, libraries, etc. These unique advantages have made Google Scholar an innovative source of information in academia (van Noorden 2014) to gather valuable demographic information (Jozaghi, 2019), and it seems unrivalled in the effective provision of scholarly documents online (Gusenbauer 2019). Jozaghi (2019) has observed the demographic representation of researchers in Google Scholar, revealing that at top-ranking universities, women and minorities are severely underrepresented in all areas of science. Nentwich et al. (2012) point towards social networks for scientists, such as ResearchGate, Academia.edu, or Mendeley. Fenner and Haak (2014) introduce unique identifiers for research like ORCID to create a single identity for a researcher to use in many different scientific digital environments.

Unlike printed documents, Bar-Ilan (2001) says, Webpages are constantly changing, being removed, or temporarily inaccessible. There is also high dynamism on academic web resources (Ortega et al. 2006; Payne and Thelwall 2007, 2008). Thus, the processes of creation and dissemination of scientific material are governed by different processes, allowing greater control (Orduna-Malea et al. 2015b). For this, Koehler (1999) defined two measures to study the dynamics of the Web: constancy (the rate of change) and permanence (the probability that pages carry the same Uniform Resources Locator, URL, over time) of Webpages. Leydesdorff and Curran (2000) looked for occurrences and co-occurrences of the triple helix (university-industry-government relationships) in Brazil and the Netherlands, showing similar patterns of growth in the number of Webpages for both countries. The usage of the web is not the same for universities and institutions. Creating, sharing, disseminating and consuming information online is completely different due to the diverse organizational cultures of these institutions. The web impact factor (WIF) was coined by Almind and Ingwersen (1997) but was criticized for its use in universities. The first works of Mike Thelwall are useful to present these criticisms to the WIF (Thelwall 2000). For him, a high value is supposed to indicate a site with more significant impact because there are many pages that link to it. It shows how academic domains tend to be underrepresented, as they are smaller, and older than commercial domains. Although these studies are quite old, considering the speed with which the web has developed in the last two decades, they serve to understand how these interdisciplinary studies focused on cybermetry have begun.

Conceptual framework

Currently, most of the population connected to Internet has access to information sources through Google. This search engine, founded in 1998, based its website management system on the discipline that we now know as cybermetrics, which at that time was incipient.

Cybermetrics is a discipline that uses quantitative methods to describe the communication methods on the Internet, web contents, its interrelations, and the consumption of information by users. Cybermetric analysis include several indicators that measure the importance of web sites based on various parameters, including the number of hosted pages, their popularity based on the number of visits received, and the number and origin of external links received.

The first Google ranking system was the algorithm called PageRank, based on cybermetric calculation of the number of links that a webpage received. This calculation has been one of the main factors in arranging rankings. After 20 years, Google's algorithm has varied, but it continues to include this and other cybermetric metrics.

Although our article does not focus on the specific study of the emerging SEO (Search Engine Optimization) discipline (Ledford 2015), we do use SEO tools and metrics as a simple instrument for data extraction to describe the proportions of STI visible websites from the results of Google. It has various techniques that are applied to measure and improve websites visibility in search engines. As a first step in the process of optimizing a website, this discipline analyzes cybermetrics' indicators to assess websites and measures their visibility and impact through search engines.

In this article, we analyze this last indicator, which is the visibility of a website using keywords, which means the position that a website occupies in the search results pages (known by the acronym SERP) when we introduce one or several keywords in the search box. The ordering of websites that appear in the SERP depends on various cybermetric factors, the most important being the number of indexed pages, the number of links received, and the users' interaction with the website.

To measure visibility in terms of positioning in the Google search engine, we analyze all the contents shown in their SERPs for a group of thematic searches applying the infosphere model.

With the term "infosphere", we do not refer to all the informative environment constituted by all the entities (Floridi 2014), but only to those visible in Google results, even though we are aware they constitute only the tip of an iceberg (Khabsa and Giles 2014). Contents are continually added, changed and/or deleted worldwide (Koehler 1999, 2002, 2004). These issues give rise to inherent technical difficulties in cataloguing and updating such a vast, diverse and complex universe (Orduna-Malea et al. 2015). The Infosphere theoretical model that we adopt has been developed to build an analytical system that encompasses the Web, and it analyses the proportions in which search engines show particular categories of websites, especially media sites, websites in the public/political sphere and social networking sites (Gonzalo-Penela 2015). The URLs that appeared in SERP can be classified as indicators of mention ordinals, i.e., depending on the position in which they have performed in results, they may be between the first and tenth on the SERP. However, we currently find two problems that affect an analysis of the results. On the one hand, nowadays it is not common for SERPs to show 10 results as the average is 8.4 results. On the other hand, there is a great variability of the absolute position of websites within the SERP, variations in the order of results of the SERPs that can occur in a very short period of time for the same query. For this reason, the application of this analysis model takes into account the appearance of URLs of different categories in search results pages of SERP without taking into account the specific order. Therefore, we use the appearance of URLs in the first SERP of the search engine as visibility metric without taking into account the exact ordering.

Methods

In order to answer our research questions, we firstly proceeded to obtain the sample object of study for SRI-IA institutions. Secondly, data sources were selected and web indicators were obtained. Thirdly a descriptive statistics analysis was conducted as in the study yielded by (Orduna-Malea and Aytac 2015a). We will focus on a single search engine: Google, as it is the most frequently used in the last decade in the majority of the countries that we have studied. Moreover, more than a decade ago Google had some competition, but today it holds an almost monopolistic position in search engines. In addition, since the application in 2007 of the 'Universal Search' model, Google's algorithm shows a large set of features useful for search engine users, such as vertical search results from image databases, videos, news, maps or books. Of course, there are limitations when using a single search engine. However, thanks to web scraping methods, a large part of the links of these resources can be extracted from the SERP and analyzed belonging to the field of informetrics. Web scraping is used to extract information from Google, and data is collected and used mainly for learning and predictive analytics (Bail 2014; Mills 2018; Siegel 2013). For instance, the descriptive statistics analysis of visibility of the conglomerates, i.e. relatively homogenous "natural" groupings in a population, composed in our case by all the URLs with visibility on the Google search results page. From this coming data, we have constructed our instrument of analysis: Infosphere conglomerates. For us, Infosphere are the set of websites that reach visibility in the search engine for a group of searches related to a specific topic, ore specifically in the case of our analysis, searches belonging to the three subjects under study, science, research and innovation. Their aggregated results would form particular info-spheres linked to each of these three themes.

Thanks to the SEO Advanced Web Ranking Cloud tool, data have been effectively collected. We believe that our data extraction methodology is correct since the alternatives are much more complex and expensive. For example, it would have been possible to capture SERP results manually from each of the selected countries, something complex and logistically costly. Moreover, a tool like ScrapeBox, that allows the use of national Internet Protocol addresses (IPs), could have been used. This alternative would have required at least one functional IP for more than one year for each of the countries, which would have had to be acquired from national suppliers. However, the possible temporary or total blocking by Google of these IPs would have required new replacement IPs, which would have increased the cost and time of data collection.

For the extraction of data, we captured the 5 first SERPs. Eleven (11) national versions of the Google search engine were used for Spain and the most representative IA Spanish-speaking countries: Argentina, Bolivia, Colombia, Chile, Ecuador, Mexico, Peru, Paraguay, Venezuela, and Uruguay. We selected for the final sample of analysis the one corresponding to the first page of results that appear in the national versions of Google when we introduce one of the 20 terms (Table 1). We have divided these 20 terms into three categories: Science, Research, and Innovation. We want to clarify that Google search adds by default the Boolean AND instead of the spaces.

The analyzed data are quantitatively provided. Moreover, these data have been treated qualitatively, as they have been categorized following a well-defined and bounded structure. The qualitative data grouped into nominal categories, thanks to our coding of the data, show the proportions of them. Also, the data collected have been organized in frequency tables that show the classes or categories of the sample (strata and groups), as well as their number and proportion. Each group shows its absolute and its relative frequency for the rest of the groups, as well as the cumulative frequency of all groups belonging to a specific state.

Findings

In our infometric analysis, we looked at the dynamic nature of this medium observing these Search Engine Result Pages over an extended period of time. This is why data collection process was carried out monthly for a total of 13 months, from July 2016 to July 2017 in each of the national search options specified above. Once all the SERPs URLs were extracted, the sample size was made up of 28,600 URLs corresponding to a total of 961 web domains. We grouped these data into three thematic Infospheres based on the keywords used for data extraction (Table 2). Of course, there is overlap in the results due to the time series and the finite number of universities.

Although we use a limited number of searches, we have tried to overcome this limitation with a 13-month time series. We believe that this series is representative of the total population. Taking this sample, we have developed a work of descriptive statistics, organizing, summarizing, and describing the data set using numerical techniques.

Each thematic Infosphere has been divided into three strata according to the nature of the websites that host the URLs. According to the adaptation of the original Infosphere model established by W.L. Bennett (Scott 2008) we can group the results that appeared in the Google SERPs in three strata. These consist on an upper layer or conventional

Table 1 The 20 keywords used for the capture

Science	Research	Innovation
Desarrollo científico (Scientific development)	*Grupo de investigación* (Research group)	*Innovación* (Innovation)
Desarrollo de la ciencia (Development of science)	I + D (R + D): Innovación y Desarrollo (Research and Development)	*Innovación científica* (Scientific innovation)
Ministerio de ciencia (Ministry of Science)	*Investigación científica* (Scientific research)	*Innovación técnica* (Technical innovation)
Publicaciones científicas (Scientific publications)	*Investigación y ciencia* (Research and science)	*Innovación tecnológica* (Technological innovation)
Ciencia y tecnología (Science and technology)	*Investigación y desarrollo* (Research and development)	*Tecnología, información e innovación* (Technology, information and innovation)
Desarrollo científico y tecnológico (Scientific and technological development)	*Proyecto científico* (Scientific project)	
Ministerio de ciencia y tecnología (Ministry of science and technology)	*Proyectos de investigación científica* (Scientific research project)	
Proyectos de ciencia y tecnología (Science and technology projects)		

Table 2 Size of the recollected SCI Infospheres

Infospheres	URLs	%
Science	11,440	40%
Research	10,335	35%
Innovation	7150	25%
Total	28,600	100%

Table 3 Strata of the infosphere applied to the total of the sample and to the results corresponding to the three topics

Layer	STI infosphere		Science infosphere		Innovation infosphere		Research infosphere	
	URLs	%	URLs	%	URLs	%	URLs	%
1. Upper layer	2244	7.84	1230	10.75	641	8.97	373	3.73
2. Middle layer	19,118	66.84	7302	63.83	4,521	63.23	7295	72.88
3. Lower layer	7238	25.30	2908	25.42	1988	27.80	2342	23.40
Total	28,600	100	11,440	100	7150	100	10,010	100

stratum composed of media websites, a middle layer composed of web portals that are relevant in response to thematic consultations raised by the search engine, and a lower layer where we group content from blogs and social networks. In our study, we consider that governmental and university websites are the most important among those web portals in response to SRI consultations. Therefore, we have grouped them into two different clusters.

In the following table, we have applied this categorization in strata to the total sample and each of the three thematic Infospheres (Science, Research, and Innovation) of which it is composed. The total of 28,600 URLs is distributed in the three areas in this way: (Table 3)

In response to the research question 1, the middle layer has the largest impact with 66.84%, more than the double of the visibility of Micro Strata and nine times more than the news websites. Although the objective of this article is not to measure the visibility of media sites, it is significant to see the low average impact they have on SERPs, especially on the research Infosphere. This fact shows that media sites generate little news about SRI and that Google's algorithms do not find relevant news to show on their SERPs. Therefore, we observe that most of the contents belong to the two lower layers in all the captured Infospheres of the 11 countries. The highest percentage of visibility of Micro Strata is in the Innovation Infosphere Stratum, to underline the importance of social networks for innovation.

As we mentioned earlier, we will focus on governmental and university websites grouped in the middle layer, which that represent 38.6% of the total sample. The governmental and university websites have the most significant impact on the Infosphere. So as to know why, we take into consideration the premises of Google for SEO: number of visits, page links, number of pages and the update of contents. Many governmental pages are frequently visited by national users to do their daily work. We can take into account students, professors, and researchers of a University that usually visit some of its pages to access their virtual campus or official email accounts.

Table 4 Results from government websites

	Google country search options																							
	Argentina		Bolivia		Chile		Colombia		Ecuador		Mexico		Peru		Paraguay		Spain		Uruguay		Venezuela		Global	
Govern-ment sites	URLs	%	URLs	%	URLs	%	URLs	%	URLs	%	URLs	%	URLs	%	URLs	%	URLs	%	URLs	%	URLs	%	URLs	%
Dominican Republic			1	0.28					1	0.31					1	0.29							3	0.08
Argentina	302	**74.38**	91	25.71	54	15.04	84	24.85	94	29.01	95	23.23	76	23.31	97	28.36	71	20.34	102	35.29	60	16.62	1126	29.19
Bolivia			110	**31.07**			5	1.48	10	3.09	3	0.73	10	3.07	10	2.92	8	2.29	6	2.08			162	4.20
Brasil											1	0.24	1	0.31			1	0.29					3	0.08
Chile	9	2.22	14	3.95	230	**64.07**	11	3.25	14	4.32	16	3.91	13	3.99	14	4.09	7	2.01	17	5.88			345	8.94
China	4	0.99	8	2.26	10	2.79	4	1.18	12	3.70	8	1.96	7	2.15	5	1.46	5	1.43	8	2.77	9	2.49	80	2.07
Colombia	2	0.49			1	0.28	135	**39.94**	2	0.62													140	3.63
Costa Rica			12	3.39	2	0.56	10	2.96	12	3.70	19	4.65	13	3.99	14	4.09	12	3.44	13	4.50			107	2.77
Ecuador									62	**19.14**													62	1.61
El Salvador	1	0.25	2	0.56																	2	0.55	5	0.13
Mexico	45	11.08	37	10.45	33	9.19	36	10.65	38	11.73	195	**47.68**	39	11.96	46	13.45	32	9.17	34	11.76	28	7.76	563	14.60
Peru	1	0.25	3	0.85	1	0.28	2	0.59	1	0.31			114	**34.97**	5	1.46	3	0.86					130	3.37
Paraguay	2	0.49	3	0.85					1	0.31			3	0.92	80	**23.39**	1	0.29	3	1.04			93	2.41
Spain	40	9.85	61	17.23	28	7.80	46	13.61	59	18.21	60	14.67	45	13.80	57	16.67	199	**57.02**	56	19.38	27	7.48	678	17.58
USA			2	0.56			3	0.89			2	0.49	1	0.31									8	0.21
Uruguay																			41	**14.19**			41	1.06
Venezuela			10	2.82			2	0.59	18	5.56	10	2.44	4	1.23	13	3.80	10	2.87	9	3.11	235	**65.10**	311	8.06
Total	406	100	354	100	359	100	338	100	324	100	409	100	326	100	342	100	349	100	289	100	361	100	3857	100

We start from the premise that websites corresponding to the state have a preponderance in each national search option. From this idea, we analyze state and university websites of all IA + Spain countries in each of Google's national search options.

In response to the research question 2, we have identified 3857 results from government websites, which represent the 13.73% of the total sample with an average of 350 government sites visible in each country sample. The highest visibility of government websites corresponds to the national search option in most cases. That is, in the search option of Google Argentina the country that most state websites rank in the first SERP is Argentina, and the same is true in most countries with few exceptions (Table 4).

On average, countries place their government websites at 39.14% on the total of them that appear in the sample of each national search engine (in bold).

The countries that own more government websites ranked in the SERP on the entire state sites detected are: Venezuela (65.1%), Spain (57,02%) and Argentina (74,38%). Conversely, the lowest number of government websites are in Uruguay (14.19%), Ecuador (19.14%), and Paraguay (23.39%).

The only countries where we find that government websites of other nationalities have more visibility than their own are Ecuador, Uruguay and Paraguay. In these countries, the highest visibility is reached by Argentinian government websites (29.01% vs. 19.14% in Ecuador, 35.29% vs. 14.19% in Uruguay and 28.36% vs. 23.39% in Paraguay).

As in government websites, we have searched for university websites (Table 5). We have identified 3358 results from university websites, 11.74% of the total sample, with an average of 305.27 university sites visible in each country sample. In response to research questions 1 and 4, we have observed that the general rule is that the highest visibility of educational websites corresponds to the national search option. That is, in the Google search option Argentina, most of the universities that reach visibility in the first SERP are Argentinian universities and so do in most of the countries. There are many other non-cybermetric factors that contribute to the Google algorithm. Especially relevant for this article is the age of the domain. In general, university domains are very old and very stable, so they are algorithmically in a better position than other organization websites to reach the top positions.

On average, countries rank their own university websites at 32.46%. However, unlike government websites where we saw three exceptions, in the case of universities we find more exceptions to this rule.

The country with most university websites rank is Spain with 1583 URL (47.14%), followed by Colombia with 490 URLs (14.59%) and Chile with 336 URLs (10.01%). As a consequence, the country with most own university websites rank is Spain with 85.19% of the websites of universities visible from Google.es. Thus, Spain is followed by Colombia with 55.70% of the websites of universities visible from Google.com.co. The country with the fewest websites is Bolivia, where only 2.39% of university websites visible from the national version of the search engine belong to this country.

In a total of 8 countries of the 11 study object, we find that websites from other nationalities have more visibility than their own national university websites. This occurs with Bolivia (own visibility 2.39%), where Spanish universities reach 60.96%, followed by Paraguay (57.71%) and Peru (49.48%).

In Table 6, we can observe how governmental websites with 3857 pages (13.49%), universities websites 3292 pages (11.51%), and Wikipedia with 2835 pages (9.91%) have a significant impact on the SRI Infosphere. Formal institutions and Wikipedia represent one-third of the entire information. Wikipedia pages have a substantial impact on many EULAC countries with 9.91% of the whole results.

Table 5 Results from university websites

Google country search options																								
Univeristy sites	Argentina		Bolivia		Chile		Colombia		Ecuador		Mexico		Peru		Paraguay		Spain		Uruguay		Venezuela		Global	
Argentina	97	30.79%	20	7.97%	17	5.03%	11	3.69%	19	6.21%	8	2.73%	10	3.46%	22	7.89%	8	2.47%	20	5.67%	14	4.49%	246	7.33%
Bolivia			6	2.39%																			6	0.18%
Chile	15	4.76%	20	7.97%	125	36.98%	22	7.38%	24	7.84%	19	6.48%	26	9.00%	19	6.81%	21	6.48%	21	5.95%	24	7.69%	336	10.01%
Colombia	31	9.84%	36	14.34%	39	11.54%	166	55.70%	41	13.40%	46	15.70%	41	14.19%	39	13.98%	1	0.31%	12	3.40%	38	12.18%	490	14.59%
Ecuador	1	0.32%	1	0.40%	1	0.30%	1	0.34%	63	20.59%	1	0.34%	1	0.35%	1	0.36%	1	0.31%	1	0.28%	1	0.32%	73	2.17%
Mexico	25	7.94%	15	5.98%	26	7.69%	26	8.72%	26	8.50%	81	27.65%	26	9.00%	13	4.66%	17	5.25%	22	6.23%	26	8.33%	303	9.02%
Peru									1	0.33%			42	14.53%	1	0.36%							44	1.31%
Paraguay															23	8.24%							23	0.68%
Spain	146	46.35%	153	60.96%	129	38.17%	72	24.16%	132	43.14%	136	46.42%	143	49.48%	161	57.71%	276	85.19%	131	37.11%	104	33.33%	1583	47.14%
Uruguay											2	0.68%							146	41.36%			148	4.41%
Venezuela					1	0.30%															105	33.65%	106	3.16%
Total	315	100%	251	100%	338	100%	298	100%	306	100%	293	100%	289	100%	279	100%	324	100%	353	100%	312	100%	3358	100%

Table 6 Results from governmental, universities, and social media websites

Governmental sites.	3857	13.49%
Academia.	7893	27.60%
University websites	3292	11.51%
Education websites	3168	11.08%
Research websites	388	1.36%
Other Academia related sites	1045	3.65%
Wikipedia.	2835	9.91%
Social Network.	2594	9.07%
Web Social.	1515	5.30%
Facebook.	517	1.81%
Pinterest.	208	0.73%
Youtube.	185	0.65%
Twitter.	95	0.33%
Academia.	66	0.23%
Linkedin.	5	0.02%
GooglePlus.	2	0.01%
Instagram.	1	0.00%
Researchgate.	0	0.00%

Table 7 Results from conglomerates by subcategories and area

	Science		Innovation	Innovation	Research	Research	Total	Total
Governmental sites.	2679	23.42%	694	9.71%	484	4.84%	3857	13.49%
Academia.	2547	22.26%	1347	18.83%	3999	39.95%	7893	27.60%
University websites	575	5.03%	377	5.20%	2340	23.38%	3292	11.51%
Education websites	1567	13.70%	749	10.48%	852	8.51%	3168	11.08%
Research websites	85	0.74%	1	0.01%	302	3.02%	388	1.36%
Other Academia related sites	320	2.80%	220	3.08%	505	5.04%	1045	3.65%
Wikipedia.	1303	11.39%	552	7.72%	980	9.79%	2835	9.91%
Social Network.	926	8.09%	797	10.45%	871	8.70%	2594	9.07%
Web Social.	440	3.85%	731	10.22%	344	3.44%	1515	5.30%
Facebook.	160	1.40%			357	3.57%	517	1.81%
Pinterest.	138	1.21%			70	0.70%	208	0.73%
Youtube.	92	0.80%			93	0.93%	185	0.65%
Twitter.	93	0.81%			2	0.02%	95	0.33%
Academia			66	0.23%			66	0.23%
Linkedin.	1	0.01%			4	0.04%	5	0.02%
GooglePlus.	2	0.02%					2	0.01%
Instagram.					1	0.01%	1	0.00%
ResearchGate							0	0.00%
	7047	61.60%	3343	46.76%	5284	52.79%	15,674	54.80%

In Table 7, we have a breakdown of each area and subcategories. The greatest weight of government websites in science information stands out with the 23.42%. This figure is significantly lower in the other two infospheres, which have replaced these sites with a greater weight of conglomerates of companies in the case of innovation, and with SRI Websites in the Research Infosphere. On the contrary, the Research infosphere has a greater number of Academia websites with the 39.95% of the results, a percentage that falls almost by half in the Science (22.26%) and Innovation (19.76%) infospheres. The group "Web Social" includes any other social network that does not belong to the most known social networks (Facebook, Twitter, etc.). There are countless social networks and interest groups that use social software to create more or less closed communities. In relation to Academic web networks, we have analyzed the visibility of the two major references: Academia.edu and Researchgate.net. Surprisingly, the visibility of both research networks is close to zero as only one URL from Academia.edu has been identified in the "scientific innovation" search sample and none from Researchgate.net. Based on our sample and these results, we can conclude that Thelwall's (2000) statement is met for academic websites, but not for academic social networks, as they are almost invisible in Google SERPs.

In this type of thematic searches, the results of the first strata have an average of 748 results with a standard deviation of 438.40. The results of the middle strata have an average of 6372.66 results with a standard deviation of 1603.59. Finally, the results of the lower strata have an average of 2412.66 results with a standard deviation of 464.05. On the contrary, when we analyze the proportions of results between countries we can observe more homogeneous data with a standard deviation. The results of searches by country of the first strata have an average of 204 results with a standard deviation of 64.97. Secondly, the results of the middle strata have an average of 1738 results with a standard deviation of 43.11, and, thirdly, the results of the lower strata have an average of 658 results with a standard deviation of 48.74.

In response to the research question 3, we have noticed that there are significant differences between types of institutions in the web in terms of proportion of government websites with respect to universities and other research centers. Government websites have an important weight in the science infosphere with 23.42%, while their weight in the innovation and research infosphere is significantly lower, with a 9.71% and a 4.84% respectively. The results that appear in the first position have their origin in different conglomerates according to thematic search group. In the case of Science searches, 43.18% of the results of the first positions come from government sites. In Innovation searches, the largest number of first positions comes from the conglomerate of virtual libraries with the 20%. Finally, in searches on Research, the largest number of first positions is on educational sites with the 24.68%.

According to the analyzed data, government websites of each country have more visibility in a national search, with few exceptions. On the other hand, in university websites we can notice more visibility on websites from Spain when compared to national universities, with exception of cases as Colombia, Uruguay, and Venezuela.

In response to research question 5, the proportions detected between the three infospheres are different. News websites show a bigger proportion in the Science infosphere with a 10.75%, while in Research infosphere the proportion falls to a 3.73%. On the opposite, results of social media websites have a similar proportion in all three infospheres, around the 25%. The role of universities stands out for its weight in the Research Infosphere with the 23.38% of the results, while the other two infospheres reach much lower proportions, with a 5.03% in Science and a 6.2% in Innovation. It is also significant that

Table 8 The Top 20 universities from 11 national versions of Google

Top20 universities		Domains	.cl	.ve	.ar	.bo	.co	.ec	.mx	.pe	.py	.uy	.es	Total
1. Escuela de Organización Industrial	Spain	eoi.es	5	8	19	10	20	14	16	22	22	17	25	178
2. Universidad Autónoma de Aguascalientes	Mexico	uaa.mx	13	14	13	13	13	13	23	13	13	13	13	154
3. Universidad de Almería	Spain	cms.ual.es	13	13	13	13	13	13	13	13	13	13	13	143
4. Universidad del Norte	Colombia	uninorte.edu.co	13	13	13	13	13	13	13	13	13	10		127
5. Universidad de Salamanca	Spain	usal.es	11	13	12	12		13	13	12	12	6	12	116
6. Universidad EAFIT	Colombia	eafit.edu.co	13	13	11	12	13	13	13	13	12			113
7. Universidad de Chile	Chile	uchile.cl	11	8	9	10	10	10	10	10	10	10	10	108
8. Universidad Autónoma Metropolitana	México	uam.mx	13	12	12	2	13	13	13	13		9	4	104
9. Universidad de Oviedo	Spain	uniovi.es	11	6	12	12		7	11	12	13	5	11	100
10. Universidad Autónoma de Madrid	Spain	uam.es	12	10	9	7		11	12	12	8	8	11	100
11. Universidad de Murcia	Spain	um.es	8	2	9	15	1	7	6	10	12	9	17	96
12. Pontificia Universidad Católica de Chile	Chile	www7.uc.cl	13	11	3	6	9	10	6	13	6	8	8	93
13. Universidad Autónoma de Barcelona	Spain	uab.cat	7	10	10	8	9	9	6	7	6	9	9	90
14. Universidad de Valencia	Spain	uv.es	6	2	7	12		7	7	4	8	8	1	62
15. Facultad de Ciencias Económicas. Universidad Nacional de Rosario	Argentina	fcecon.unr.edu.ar	5	5	11	8	9	2	4	2	6	5	4	61
16. Institución Universitaria Antonio José Camacho	Colombia	uniajc.edu.co	5	8	3	3	13	8	8	6	4			58
17. Universidad de Málaga	Spain	uma.es	3	1	6	5		3	2	4	4	4	7	39
18. Universidad La Gran Colombia	Colombia	ugca.edu.co	5	2	3	3	6	3	5	6	6			39
19. Universidad Carlos III de Madrid	Spain	uc3m.es	3	1	5	5		2	3	4	5	3	3	34
20. Universidad Politécnica de Valencia	Spain	upv.es	1	1	1	4	1	5		2	5	3	7	30

universities only have top positions in Research searches with the 15.58% of the total content positioned at the first position.

Table 8 shows a specific map of the Top 20 universities from 11 national versions of the Google search engine from the most representative Ibero-American countries: Chile (.cl), Venezuela (.ve), Argentina (.ar), Bolivia (.bo), Colombia (.co), Ecuador (.ec), Mexico (.mx), Peru (.pe), Paraguay (.py), Uruguay (.uy), and Spain (.es).

Web relationships heavily depend on linguistic, cultural, and proximity factors (Vaughan 2006). For this reason in the Top 20 positions, we find 11 Spanish universities. This show the important presence of this country in the LAC panorama for language and culture, influenced by socio-political factors and socio-economic characteristics (Adams et al. 2014, Finardi and Buratti 2016, Chinchilla et al. 2018). The institution named "Escuela de Organización Industrial" that appears in the first position is a Business School and not a university although it may have cooperation agreements with some public or private universities. The second, the third, and the fourth have a similar impact and structure. They are small-medium universities compared to other universities in these countries, not well known, but high ranked in different national versions of Google. Results are very similar to the rest of the Top 20 ranked universities, with few exceptions (position 7, 8, 10, 12, 13, 19) where the most prestigious and well-known institutions appear in different national versions of Google. Mexico and Spain are the only countries where a local university appears as the first result in their national version of Google. It is surprising how a small Spanish university from 1955, located in Madrid, is the top-ranked. It is the best ranked in the seven national versions of Google. Table 8 includes subdomains because, despite being hosted under the same domain, they can be part of to completely different websites both thematically and technologically, and belong to specific faculties or prominent centers within the university.

We consider that these results are the consequences of two causes. Firstly, websites in the first twenty positions have a large number of links with other important Webpages of institutions, which increases the number of possibilities of appearing in the first places. Moreover, these universities know how Infospheres work and how to be positioned at the top.

Discussion

In this paper, we have observed the impact that institutional websites have on Ibero-American countries. We noticed that the composition of Google results varies significantly when we search about Science, Innovation or Research. From a socioeconomic perspective, the composition and evolution of content affects its consumption in different countries according to varying social, economic and political issues (Cortés-Sánchez 2019, Orduna-Malea et al. 2015b). Positioning their own institution in the first results is the goal of each organization with an international vision. Through the analysis of the ranking of these websites, limited to a specific period (thirteen months), we determined if web positioning is a strategy for educational and research institutions.

In our research, we have analyzed how the data obtained are an indicator of digital practices displayed both by users that do the searches and institutions that manage websites, according to Search Engine Optimization (SEO) criteria. The pages that appear more frequently in the first positions in IA countries are not national but institutions from other countries, as in the university Top 20 ranking. In addition, we have observed that governmental and educational institutions are more visible than research institutions. Many times, foreign institutions have more impact than national institutions. Although social networks are not too popular for this type of institutions, they are gaining positions in the last period.

The role that university websites play in this emerging scenario is strategical. It is important to map this scenario to understand the type of representation that each country has in the international arena, to observe the relations of academic dependency of small and developing countries (Foladori 2006; Kay and Shapira 2009, 2011), and to understand how scientific collaboration will be fostered in future scenarios. We trust that universities, like other institutions, have an international purpose to attract students and researchers and need to improve their ability to be visible in the Infosphere. IA countries represent a specific and emerging scientific context. Shared language and similar culture favor mobility of students, researchers, and professors among them. The impact of Spanish language is evident, so none of the universities in the United States or the United Kingdom have been identified in the sample, despite being world references. The lack of content in Spanish makes them invisible to Google's algorithms when searching in this language. Many postgraduate students and researchers from IA countries are characterized by high mobility to European countries, mostly Spain, and the United States. International scientific collaboration rates of IA countries show significant networking practices that could be reflected on SRI websites Infosphere. We have identified the role that universities play in the SRI IA and Spain Infosphere. Thanks to it, we can present to the reader how an Infosphere is built and positioned.

Limitations

There are a number of limitations to this study inherent in the design of the data collection method. One of the main limitations has been the absence of researches on the subject since, despite previous rankings of universities from different countries (Dominic and Jati 2010; Kaur et al. 2016; Olaleye et al. 2018), we lack previous literature based on Google visibility studies of universities using thematic groups of keywords.

These studies are carried out by SEO professionals for companies, but they are not published in academic journals due to their professional and private nature.

On the other hand, the SEO tool used in the process of extraction of results is hosted in the cloud and provides results that simulate the searched language and geolocation in a server of the selected country. However, we could not contrast the results with other similar tools or with manual searches using national IPs. All measures based on web indicators should be taken with some caution (Orduna-Malea and Aytac 2015a), in particular, those obtained from a search engine like Google, due to the variability of the data (Wilkinson and Thelwall 2013) and the geolocalization of results. For this reason, a second capture in parallel had allowed us to measure possible deviations from our main tool and correct our conclusions.

We also think that the size of the sample studied can be seen as a limitation. It should be noted that, although 28,600 URLs have been analyzed, they come from eleven national search engines, leaving national samples at only 2600 URLs per country.

The keywords used for defining the infospheres can be interpreted ambiguously. For examples in the case of "Scientific project" and "Projects of science and technology", we understand that research is developed in the form of a scientific project, while "Scientific project" can include other realities outside research. The Innovation group includes words from the same lexical family, as well as the Science group. In the case of Research there is an exception, that is, not all belong to the same lexical family.

Another limitation has been to measure the visibility of the sites using the number of URLs that appeared in the first positions of the SERP without taking into account their exact position. Adding the CTR (Click Through Rate) or the probability of clicks according to its position would have distorted the proportions and new indicators would have to be calculated.

In summary, search engines keep huge collections of web documents and sometimes they may exhibit unpredictable and inconsistent behaviour. Researchers need to take into account these characteristics of search engines (Uyar 2009).

Conclusions

Due to the amount of data collected from multiple search options, in our paper we have worked on visibility according to the appearance (or not) of university websites on the SERP for the selected keywords, without considering their exact position on the results website. We have only analyzed the sites that appear on the first position. It is fundamental to appear in these high positions since they increase the possibility of having more clicks which can in turn become users' visits. In a deeper analysis, considering the position occupied by the universities, we could calculate the probability of accumulated clicks of universities against the results of other conglomerates.

Unless academic institutions have a person responsible for SEO to control the Google positioning, webmasters and website editors only upload information to the websites. Therefore, the Google algorithm determines the sites that appear in the first SERP, firstly based on origin, and then based on multiple ranking factors that include quality, topology, and freshness of its content. The fact that institutions just upload content and do not care about the visibility of their websites is reflected in the lack of a responsible web optimization and the lack of free web analytics systems such as Google Analytics. Thus, it is impossible for them to determine their position in Google, if they appear or not in relevant searches and which contents are the most relevant according to the number of visits. Regarding question 6 about what is the role played by university websites in this scenario, we can say that universities are concerned about their position in the ranking of universities based on impact and academic productivity criteria, but do not take into account the importance of having an optimized website appear in high positions on the SERP.

To conclude, we point out that the connectivity of a researcher in a scientific network is more important than the quality of research. We claim that Google results act as a gateway to these networks through its results, limiting the visibility of national websites from which search is performed.

Acknowledgements This paper has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 693781 – "Giving focus to the Cultural, Scientific and Social Dimension of EU – CELAC Relations".

References

Ackland, R., Fry, J., & Schroeder, R. (2007). Scoping the online visibility of e-research by means of e-research tools. In Proceedings of the 3rd international e-social science conference.

Adams, J., Gurney, K., Hook, D., & Leydesdorff, L. (2014). International collaboration clusters in Africa. *Scientometrics, 98,* 547–556. https://doi.org/10.1007/s11192-013-1060-2.

Almind, T. C., & Ingwersen, P. (1997). Informetric analyses on the world wide web: Methodological approaches to 'webometrics'. *Journal of documentation, 53*(4), 404–426.

Arslan, M. L., & Seker, S. E. (2014). Web based reputation index of Turkish universities. *International Journal of e-Education, e-Business, e-Management and e-Learning, 4*(3), 197–202.

Bail, C. (2014). The cultural environment: Measuring culture with big data. *Theory and Society, 43*(3–4), 465–482.

Bar-Ilan, J. (2001). Data collection methods on the web for infometric purposes—a review and analysis. *Scientometrics, 50*(1), 7–32.

Beaulieu, A. (2001). Voxels in the brain: Neuroscience, informatics and changing notions of objectivity. *Social Studies of Science, 31*(5), 635–680.

Belli, S., & Baltà, J. (2019). Stocktaking scientific publication on bi-regional collaboration between Europe 28 and Latin America and the Caribbean. *Scientometrics, 121*(3), 1447–1480.

Berman, D. (2004). At bell labs, hard times take toll on pure science. *Wall Street., 23,* 1–12.

Callon, M., & Law, J. (2004). Introduction: Absence—presence, circulation, and encountering in complex space. *Environment and Planning D: Society and Space., 22,* 3–11.

Chinchilla, Z., Miguel, S., Perianes-Rodríguez, A., & Sugimoto, C. (2018). Dependencies and autonomy in research performance: Examining nanoscience and nanotechnology in emerging countries. *Scientometrics, 115*(3), 1485–1504.

Codina, L., Gonzalo-Penela, C., Pedraza, R., & Rovira, C. (2017). *Posicionamiento web y medios de comunicación: ciclo de vida de una campaña y factores SEO*. Barcelona: Departamento de Comunicación. Serie Editorial DigiDoc.

Cortés-Sánchez, J. D. (2019). Innovation in Latin America through the lens of bibliometrics: Crammed and fading away. *Scientometrics, 121*(2), 869–895.

Dominic, P. D. D., & Jati, H. (2010). Evaluation method of Malaysian university website: Quality website using hybrid method). *2010 International symposium in information technology* (ITSim (Vol. 1, pp. 1–6). IEEE.

Drèze, X., & Zufryden, F. (2004). Measurement of online visibility and its impact on internet traffic. *Journal of interactive marketing, 18*(1), 20–37.

Fenner, M., & Haak, L. (2014). Unique identifiers for researchers. In S. Bartling & S. Friesike (Eds.), *Opening science* (pp. 293–296). Cham: Springer.

Finardi, U., & Buratti, A. (2016). Scientific collaboration framework of BRICS countries: An analysis of international co-authorship. *Scientometrics, 109,* 433–446. https://doi.org/10.1007/s11192-016-1927-0.

Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford: Oxford University Press.

Foladori, G. (2006). Nanotechnology in Latin America at the crossroads. *Nanotechnology Law & Business, 3,* 205–216.

Gonzalo-Penela, C. (2015). *Posicionamiento web y dinámicas de información en motores de búsqueda: propuestas de análisis y estudio comparativo de visibilidad de contenidos digitales en el caso de procesos electorales* (Doctoral dissertation, Universitat Pompeu Fabra).

Gusenbauer, M. (2019). Google Scholar to overshadow them all? comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics, 118*(1), 177–214.

Hine, C. (2002). Cyberscience and social boundaries: The implications of laboratory talk on the internet. *Sociological Research Online, 7*(2), 1–16.

Jozaghi, E. (2019). A new innovative method to measure the demographic representation of scientists via Google Scholar. *Methodological Innovations, 12*(3), 1–9.

Kaur, S., Kaur, K., & Kaur, P. (2016). An empirical performance evaluation of universities website. *International Journal of Computer Applications, 146*(15), 53–62.

Kay, L., & Shapira, P. (2009). Developing nanotechnology in Latin America. *Journal of Nanoparticle Research, 11,* 259–278.

Kay, L., & Shapira, P. (2011). The potential of nanotechnology for equitable economic development: The case of Brazil. *Nanotechnology and the challenges of equitu, equality, and development* (pp. 309–329). New York: Springer.

Khabsa, M., & Giles, C. L. (2014). The number of scholarly documents on the public web. *PLoS ONE, 9*(5), e93949.

Koehler, W. (1999). An analysis of web page and web site constancy and permanence. *Journal of the American Society for Information Science, 50*(2), 162–180.

Koehler, W. (2002). Web page change and persistence—a four-year longitudinal study. *Journal of the American Society for Information Science and Technology, 53*(2), 162–171.

Koehler, W. (2004). A longitudinal study of web pages continued a consideration of document persistence.*Information Research*, 9(2). http://informationr.net/ir/9-2/paper174.html. Accessed 01 Feb 2020.

Ledford, J. L. (2015). *Search engine optimization Bible*. London: John Wiley & Sons.

Leydesdorff, L., & Curran, M. (2000). Mapping university-industry-government relations on the Internet: The construction of indicators for a knowledge-based economy. *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics*, *4*, 1–17.

Mills, K. A. (2018). What are the threats and potentials of big data for qualitative research? *Qualitative Research, 18*(6), 591–603.

Minniti, S., Santoro, V., & Belli, S. (2018). Mapping the development of open access in Latin America and Caribbean countries. An analysis of web of science core collection and SciELO citation index (2005–2017). *Scientometrics, 117*(3), 1905–1930.

Nentwich, M. (2003). *Cyberscience: Research in the age of the internet* (pp. 479–489). Vienna: Austrian Academy of Sciences Press.

Nentwich, M., Konig, R., & König, R. (2012). *Cyberscience 2.0: Research in the age of digital social networks* (Vol. 11). Frankfurt, New York: CampusVerlag.

Olaleye, S. A., Sanusi, I. T., Ukpabi, D., & Okunoye, A. (2018). Evaluation of Nigeria universities websites quality: A comparative analysis. *Libryra Philosophy and Practice*, *1717*, 1–14.

Orduna-Malea, E., & Aytac, S. (2015). Revealing the online network between university and industry: The case of Turkey. *Scientometrics, 105*(3), 1849–1866.

Orduna-Malea, E., et al. (2015). Methods for estimating the size of Google Scholar. *Scientometrics, 104*(3), 931–949.

Ortega, J. L., Aguillo, I., & Prieto, J. A. (2006). Longitudinal study of content and elements in the scientific web environment. *Journal of Information Science, 32*(4), 344–351.

Payne, N., & Thelwall, M. (2007). A longitudinal study of academic webs: Growth and stabilisation. *Scientometrics, 71*(3), 523–539.

Payne, N., & Thelwall, M. (2008). Do academic link types change over time? *Journal of Documentation, 64*(5), 707–720.

Scott, D. T. (2008). *Tempests of the blogosphere: Presidential campaign stories that failed to ignite mainstream media* (pp. 271–300). Digital media and democracy: Tactics in hard times.

Siegel, E. (2013). *Predictive analytics: The power to predict who will click, buy, lie, or die*. Hoboken: John Wiley and Sons.

Thelwall, M. (2000). Web impact factors and search engine coverage. *Journal of documentation, 56*(2), 185–189.

Thelwall, M. (2002). A comparison of sources of links for academic web impact factor calculations. *Journal of Documentation, 58*(1), 66–78.

Thelwall, M., & Harries, G. (2004). Do the web sites of higher rated scholars have significantly more online impact? *Journal of the American Society for Information Science and Technology, 55*(2), 149–159.

Uyar, A. (2009). Investigation of the accuracy of search engine hit counts. *Journal of information science, 35*(4), 469–480.

van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature, 512,* 126–129.

Vaughan, L. (2006). Visualizing linguistic and cultural differences using web co-link data. *Journal of the American Society for Information Science and Technology, 57*(9), 1178–1193.

Wilkinson, D., & Thelwall, M. (2013). Search markets and search results: The case of Bing. *Library & Information Science Research, 35*(4), 318–325.