



Perception and trends towards research and technological development in the avocado value chain using data science tools¹

Jeimmy Cáceres-Zambrano^{2*} , Joaquin Guillermo Ramírez-Gil³ , Dursun Barrios² 

10.1590/0034-737X2024710020

ABSTRACT

Systematic analysis of scientific, industrial, and commercial information and the application of data science (DS) are alternatives to learn about trends and advances in value chains such as avocado. The objective of this study was to review current trends in research, technological development, and perception about the avocado subsector under tropical conditions using DS tools. A bibliometric analysis (BA) of scientific articles, technological developments, and patents in different sources of web information was implemented. Additionally, search algorithms and DS were designed to perform an analysis of trends from social media platforms in the avocado value chain. The BA identified patterns associated with the avocado chain in Colombia, highlighting the increase in scientific-technical information, the grouping by work networks, the bias associated with departments and commercial varieties such as the high volume of works in Antioquia and cv. Hass, and the few high-value alternatives for the chain such as patents, industrial secrets, etc. Trends in social networks were linked to international markets, added value, and environmental aspects. Likewise, a disconnection was found between the needs of the avocado value chain and the existing offers. This study contributes as a competitive strategy for the local avocado subsector and generates a country-adjusted research agenda.

Keywords: bibliometric analysis; social media platforms; knowledge management; innovation; *Persea americana*.

INTRODUCTION

Avocado (*Persea americana* Mill.) has desirable characteristics for the food, cosmetic and pharmaceutical industry worldwide (Araújo *et al.*, 2018). The international demand for this fruit has led to an increase in crop production and planted area (ITC, 2021). In Latin America, Mexico (48.0%), Peru (16.2%), Chile (4.2%) and Colombia (3.0%) account for the highest participation in the export volume of avocado in the region (ITC, 2021). In Colombia, the crop-planted area expanded very fast in the last decade (FAO, 2021)

Reports have shown that the avocado production

chain in Colombia uses foreign technologies and requires technological advances at the local level (Arias Bustos & Moors, 2018; Quintero-Ramírez *et al.*, 2019; Carranza *et al.*, 2020). The improvement in the integrated agronomic management of this crop requires research that allows regional decision-making, considering specific climatic, edaphic and topographic factors (Ramírez-Gil & Morales-Osorio, 2021; Sonnino & Ruane, 2013). Ramírez-Gil (2018) affirms that local information enables an integrated management of the avocado wilt complex disease, the most limiting pathology of the crop.

Submitted on March 16th, 2022 and accepted on January 05th, 2024.

¹ This article is the first chapter of the master thesis "Level of technological adoption in agribusiness producing avocado (*Persea americana* Mill.) and its associated factors".

² Grupo de investigación Biogénesis, Departamento de Desarrollo Rural y Agroalimentario, Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, Bogotá, Colombia. jacacerez@unal.edu.co; dbarrio@unal.edu.co

³ Grupo de Investigación Biogénesis, Laboratorio de Agrocomputación y Análisis Epidemiológico, Center of Excellence in Scientific Computing-CoE-SciCo, Departamento de Agronomía, Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, Bogotá, Colombia. jgramireg@unal.edu.co

*Corresponding author: jacacerez@unal.edu.co

Knowledge and innovation in the agricultural sector allow us to face environmental, productive and fruit quality challenges in avocado production systems (Ramírez-Gil *et al.*, 2019). The joint work of academia, institutions, and farmers based on technology transfer and adoption is fundamental (Vinholis *et al.*, 2021). In agriculture, technology plays a key role in the competitiveness of the sector, in addition to aspects such as diversity, sustainable management of the production system, knowledge management and product quality assurance (Aweke *et al.*, 2021, Cáceres-Zambrano *et al.*, 2022). The technology must be transversal to the value chain and its analyses; this is why connecting digital technologies in agriculture allows the analysis to be expanded (Klerkx *et al.*, 2010).

Bibliometric analysis is a popular, rigorous and robust quantitative method widely used in scientific research to assess the literature and analyzing large volumes of data (Donthu *et al.*, 2021). The use of bibliometric techniques in the context of the avocado chain can systematically evaluate a vast array of published works to identify key research themes, influential authors, and collaborative networks, among others (Donthu *et al.*, 2021). Through this type of analysis, it has been possible to synthesize and find patterns in multiple topics. Within these are the food supply chains and other areas of interest of the agro-fruit sector such as avocado (Tirado-Kulieva *et al.*, 2022; Kumar *et al.*, 2022; Zakaria *et al.*, 2022)

Through bibliometric analyses, we can shed light on the most prominent technological advancements and knowledge gaps in the domain, providing researchers, policymakers, and industry stakeholders with critical insights to guide their efforts and investments. Additionally, there are currently multiple software that have facilitated bibliometric analyzes and have made them very popular in different fields of scientific research (Dervis, 2019; Moral-Muñoz *et al.*, 2020).

However, while academic research is essential to shape the trajectory of the avocado industry, it is equally vital to understand the perceptions and attitudes of the broader public. Social media platforms, including Twitter, Google, and YouTube, serve as rich sources of user-generated content, reflecting real-time opinions, experiences, and discussions surrounding technological developments in the avocado sector (Wright *et al.*, 2010; Fan & Gordon, 2014; He *et al.*, 2015). Integrating social media perceptions with bibliometric analysis offers a holistic view of how technology is perceived, embraced, or challenged by various

stakeholders, ranging from consumers and producers to researchers and enthusiasts in the avocado value chain.

Bibliometric analysis and the extraction, systematization and recognition of patterns associated with information from social media platforms require advanced analysis techniques such as data science (DS) (He *et al.*, 2015). DS is an emerging field science with the aim of extracting knowledge from multiple types of data (van der Aalst, 2016). DS uses multiple algorithms, processes, and system to make friendly use of advanced techniques from statistics, mathematics, computer science, artificial intelligence, among others (van der Aalst, 2016). DS provides a broad perspective on trends in search engines and social networks (Jun *et al.*, 2018; Khan *et al.*, 2021). The latter, in turn, are pathways to meet offers and demands of a sector, because some are platforms for dissemination of opinions, events, and commercial information (Muhammad *et al.*, 2021). The methodology has become relevant due to the increase in the use of social networks and information due to the Covid-19 contingency (Rakshit *et al.*, 2021).

Based on the information above, the hypothesis of this study is that systematic search for information and application of data analytics tools allow gaining knowledge of advances and trends in research and development of the avocado subsector, as well as perception towards the value chain. Thus, the objective of this study was to review the trends in research and technological development in the avocado subsector using data science tools, to broaden the perspective on the offers and demands of this agribusiness in the geographical area studied. To achieve this, we propose a novel methodology in which data science tools and bibliometric analysis are integrated to manage and identify technological offers and demands developed in the avocado production system for the tropical condition of Colombia.

MATERIALS AND METHODS

Methodological scheme

Our work presented a methodological scheme with three phases: (i) bibliometric analysis, (ii) trends in social media platforms, and (iii) gaps between technological offers and demands in avocado cultivation (Figure 1). (i) Bibliometric Analysis: In this phase, a comprehensive bibliometric analysis was performed to identify and review existing academic literature, research articles, and relevant publications related to avocado cultivation and its technological aspects utilizing various academic databases and citation indexes. (ii) Trends in Social Media

Platforms: In the second phase, a thorough examination of social media platforms was performed to identify emerging trends, discussions, and conversations related to avocado crop technology. Various social media platforms such as Twitter, Facebook groups, you tube and relevant online forums were analyzed to gauge the sentiment, interests, and opinions of stakeholders, including farmers, researchers, industry experts, and consumers. Phase (iii) Gaps between technological offers and demands in avocado culture: The third phase involves an assessment of the gaps between technological offers and demands in avocado crops under tropical conditions. Our approach was based on comparing existing technological solutions and advances in the academic literature with the needs and challenges expressed by stakeholders in social media, the political sector, and potential gaps were identified (Figure 1).

Bibliometric analysis

The technological offer was identified from research trends in the Colombian avocado subsector presented in scientific, academic, technical, and industrial publications, as tangible final product of research and innovation processes (Flórez *et al.*, 2012). A search was carried out in scientific and academic databases, institutional repositories, and national and international online catalogues (Table 1). We used keywords and connectors in English and Spanish with simple or compound words: ‘Avocado’, ‘Persea americana’, ‘aguacate’ AND ‘Colombia’ and ‘Colombian’. The search was conducted from 01-01-2000 and 03-30-2021 and only documents whose results were evaluated in Colombia were included.

The search included repositories of 128 Higher Education Institutions registered in the National Higher Education Information System - SNIES in September 2020 (Ministerio de Educación Nacional, 2020). Forty-four institutions not related to the study area and 127 without an online catalogue were omitted. Chapters of books resulting from research related to avocado and undergraduate works filtered by title words were included, for those that had no keywords included. Thesis advisors and co-advisors were considered co-authors due to their importance in the academic and scientific community, in order to enrich the co-authorship analysis.

Keywords and author names were collected and refined with the Mendeley (2020) software (™). Through Mendeley software, the first data purification was achieved, eliminating duplicate references in the library. We recorded the systematic search process for audit purposes (Page *et al.*, 2021). The names are the second filter for the removal of duplicate documents. A thesaurus with 1,525 authors was generated, homonymy was investigated, and surname spelling was examined. Those authors with more than 3 documents in the database were considered. The keyword co-occurrence analysis used a thesaurus of 2198 words to check spelling, words in the same area, synonyms, and foreign words. A minimum of 10 repetitions per word in the documents were selected. The words avocado, agriculture, and Colombia were excluded due to their generality. We use full counting, in which each author or keyword has the same weight, regardless of the number of authors or keywords in the publication.

The bibliometric network analysis of co-authorship and

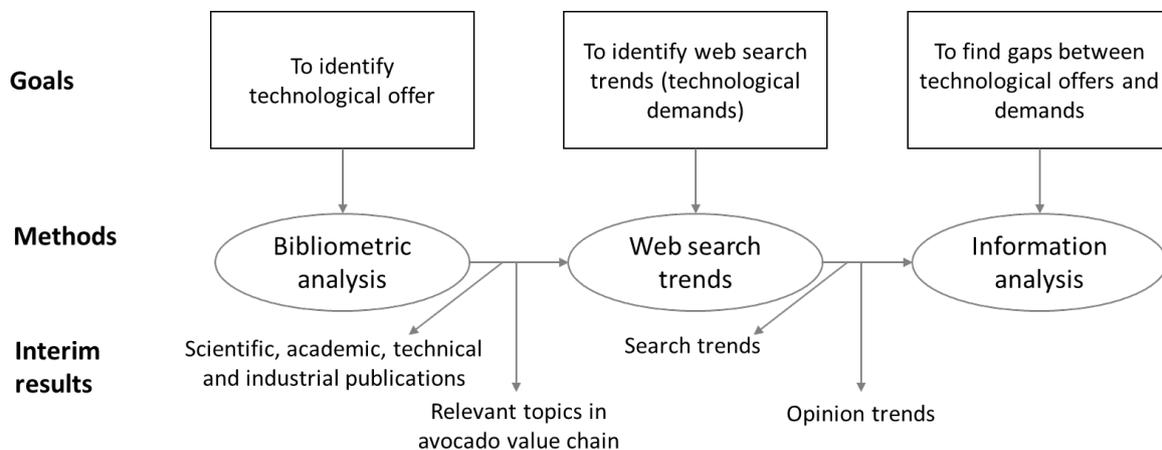


Figure 1: Scheme of our methodological approach associated with the goals, methods and results.

Table 1: Database searched uses to bibliometric analysis

	Platform
Scientific Electronic Library Online - SCIELO	Literatura Latinoamericana y del Caribe en Ciencias de la Salud - Lilacs
Fundación Dialnet	FAO Document Repository
World Wide Science	Instituto Humbolt
Academic Search Complete	Centro Internacional de Agricultura Tropical
System for Agricultural Science and Technology - AGRIS	Biblioteca Agropecuaria de Colombia
SpringerLink	SIEMBRA
Scopus	Red Colombiana de Información Científica
Science Direct	Academia Colombiana de Ciencias Exactas Físicas Y Naturales
Web of Science	Banco de La República de Colombia
Open Access Theses And Dissertations	Instituto Interamericano de Cooperación para la Agricultura IICA
Directory of Open Access Journals - DOAJ	Red Investigadores de Economía
Wiley Online Library	Revista de La Academia Colombiana de Ciencias Exactas, Físicas Y Naturales
Science.gov	Legado Renata (SNAAC)
Bielefeld Academic Search Engine - BASE	Buscador MinCiencias - Prototipos Industriales
Google académico	Publindex
Red de Revistas Científicas de América Latina y el Caribe - Redalyc	Repositorio Ministerio de Ciencias y Tecnología (Kujane)
Alianza de Servicios de Información Agropecuaria - SIDALC	Colombia Productiva
LA Referencia	Portal de Revistas de la Universidad Nacional de Colombia
Red de Repositorios Latinoamericanos	

Search date: April 2021.

keywords was carried out with the free software VosViewer (<https://www.vosviewer.com>). The distances in the generated technological maps were based on the similarity visualization technique (van Eck & Waltman, 2014). From the technological maps, we calculated the density metrics, which connect the number of existing relationships with possible associations, and centrality. Centrality measures the intensity of existing links between clusters in UCINET software (Borgatti *et al.*, 2002; Callon *et al.*, 1991). The iterative review of the information made it possible to eliminate duplicate research and to be sure that the documents linked to the analysis fit the search object.

To find the search strings by geographical location, we used the free Python programming language, in the interactive environment Google Colab (Kluyver *et al.*, 2016), from which functions were implemented to perform text mining using Numpy libraries (Harris *et al.*, 2020), Spacy (Honni-bal & Montani, 2017), and Pandas (McKinney, 2010).

Using this programming language increased the effi-

ciency of the research process since it was not necessary to review each document. At the end of the code run process, a localization verification was carried out, however, the time spent was less than what would have been needed without the use of Python.

Web search trends

Social networks are a source of exchange of experiences and information. For this, to identify the trends in the network, we entered the term “Aguacate” in Google Trends, a web tool that provides information using the Google search engine (Google, 2021a). This system of algorithms captures data associated with multiple fields of knowledge, filtering information by keywords, and determining the relative importance of a term in each time interval and a specific geographical area, linking topics according to trends in the web and the type of search (Jun *et al.*, 2018). The search for trends focused on Colombia, with records between 01-01-2010 and 06-30-2021. The

total Google Trends information was obtained by “Web Search” and the information was disaggregated by search type, Google images, Google news, Google Shopping and YouTube (Google, 2021b).

The analysis of opinion trends in the avocado value chain on Twitter was carried out in Python with specific functions for this purpose using the Pandas libraries (McKinney, 2010), Snsrape, and Itertools (JustAnotherArchivist, 2020; Tomlins, 2020). The search equation was “(aguacate OR avocado) AND (cultivo OR crop OR agricultura OR agriculture)”, the search window ranged from 2011-01-01 to 2021-06-30. The search coordinates were set at 4°4’58.8”N and 72°57’0.0”W for a radius of 1300 km and, later, it was filtered for Colombia. Accessing information from social networks allows the analysis framework to be expanded and access to information from a larger population than would be achieved through traditional instruments.

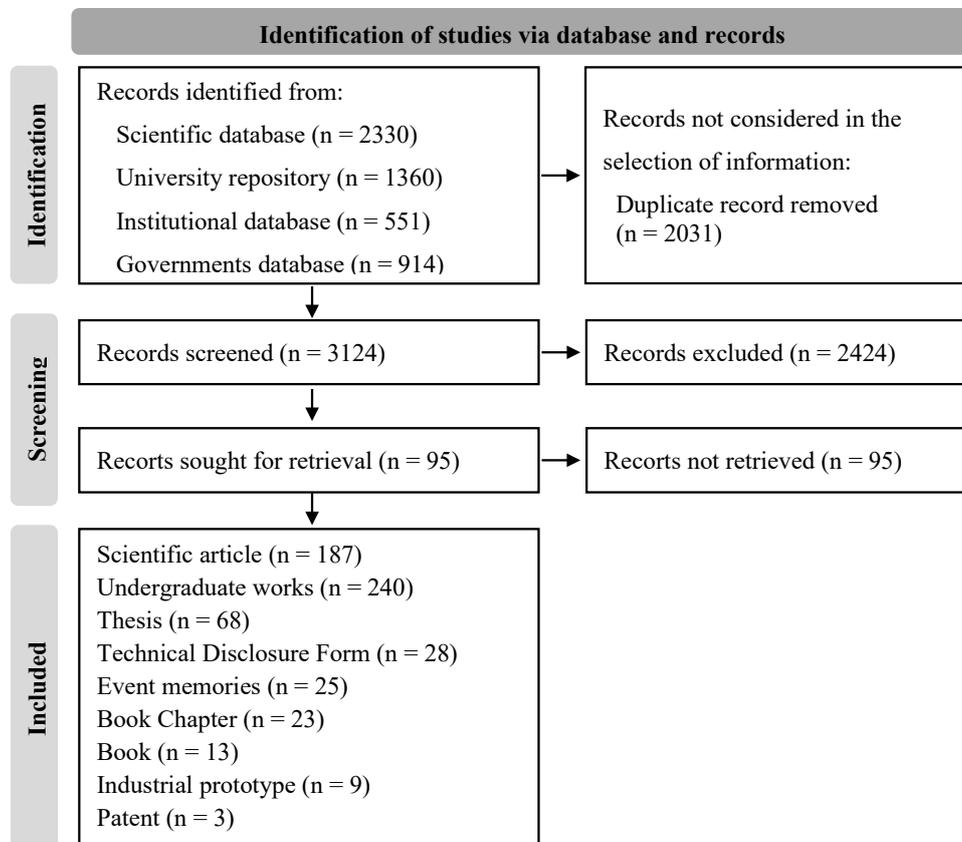
RESULTS

Bibliometric analysis

Auditing and debugging of the information allowed one to go from 5155 to 605 records, thus providing a correct filtering of the information (Figure 2).

An increase in publications related to avocados was observed in 2010 under tropical conditions, the case of Colombia, with an average annual increase of 20%, reaching 96 titles in 2020. In this sense, the Undergraduate Works aimed at the transformation of fruit and exporting to international markets stand out. Some work has similar research goals, in which information is replicated.

Of the total number of publications identified, 39.7% correspond to undergraduate works on technical and professional training and specialization. Scientific papers corresponded to 31%, with documents in Spanish (53%) and English (47%). The Master and PhD theses corresponded



Source: Author’s creation adapted from Page *et al.*, 2021.

Figure 2: Flow diagram for the systematic review and debugging of the records.

to 11.2% of the total titles, with a higher proportion of Master's theses. Technology transfer documents corresponded to 4.1%, which included bulletins, brochures, consulting documents, leaflets, reports, manuals, standards, protocols, and videos. The search identified to some lesser extent presentations at scientific events (congresses, symposiums, seminars) (4.6%), book chapters (3.8%), books (3.6%), industrial prototypes (1.5%) and patents (0.5%) (Figure 3).

The analysis identified three patents on avocado products related to cosmetic use, tree girdling, and controlled atmospheres and hormones. Additionally, 9 industrial prototypes were found for the harvest and post-harvest stages, four of these being requested by the "Grupo de investigación, Diseño y Desarrollo Aplicado" (GRINDDA) of the "Servicio Nacional de Aprendizaje" (SENA). The institutions with the greatest participation in the bibliographic results were the Universidad Nacional de Colombia (120), Corporación Colombiana de Investigación Agropecuaria - AGROSAVIA (90), Universidad Nacional Abierta y a Distancia (42), Universidad EAFIT (16), Universidad Tecnológica de Pereira (13), Universidad de Antioquia (12), Universidad de los Andes (12), Universidad Piloto de Colombia (12) and the Institución Universitaria ESUMER (10).

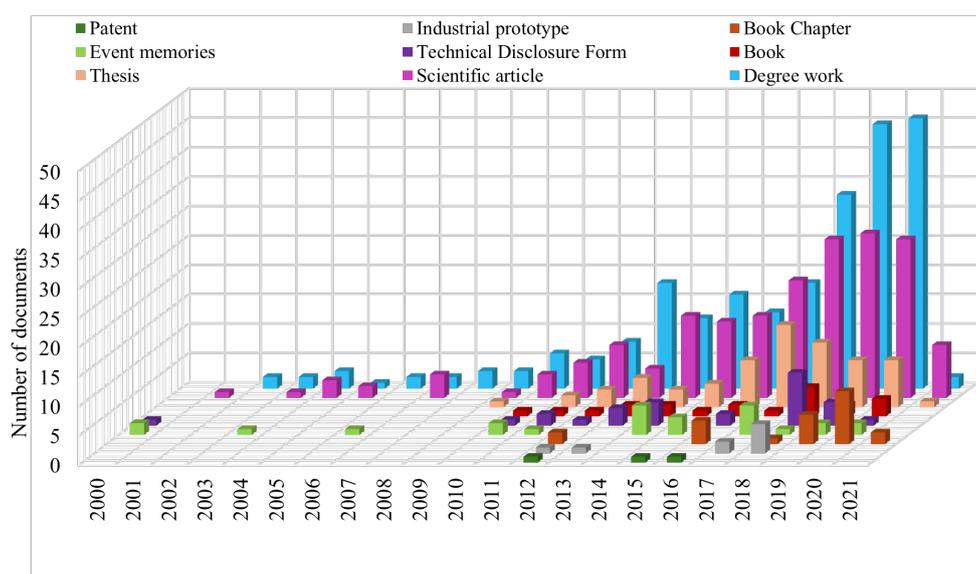
The department with the highest number of records was Antioquia, followed by Tolima and Valle del Cauca (Table 2). In total, 373 documents report the genetic material with which the research was carried out, some with evaluations in more than one variety. The Hass variety predominated with 54.2% of the total documents, followed

by Lorena (6.0%), Choquette (3.6%), Fuerte (2.1%), Reed (1.8%), Booth (1.6%), Criollo (1.5%), Papelillo (1.5%), Santana (1.5%), Trinidad (0.8%), Trapp (0.8%) and Col-linred (0,3%). The varieties Duque, Carmero, Gwen and Waldin presented only one report.

According to the co-authorship analysis, it is observed that there is collaborative work between researchers in the avocado subsector, since 7 clusters connected to each other were generated (Figure 4). The cluster with the largest number of connections (4) was made up of researchers Joaquín Guillermo Ramírez-Gil (Universidad Nacional de Colombia, Bogotá), Juan Gonzalo Morales-Osorio (Universidad Nacional de Colombia, Medellín), Juan Cámilo Henao-Rojas (AGROSAVIA) and Nelson Walter Osorio (Universidad Nacional de Colombia, Medellín). These authors mainly address issues of production, climate, biomodeling, fruit quality, new technologies (agriculture 4.0) and integrated management of pests and diseases. The centrality and density metrics show that coauthoring networks are weak.

The co-authorship network analysis found that 44.6% of the documents collected correspond to the generation of new knowledge, 25.3% to revisions, 18.0% to validation of methodologies with varieties not yet studied or in regions not previously evaluated. Keyword analysis identified five clusters that contain the main topics being investigated in Colombia.

The research topics found in this study support the creation of a broad research agenda. The words with the



Source: Authors.

Figure 3: Bibliographic records about avocado in Colombia by document type and publication year from 2000 to 2021.

Table 2: Bibliographic records related to avocado in Colombia by document type, department, and locality

Dept.	Municipalities	Number of documents								TOTAL	
		Undergraduate works	Scientific article	Thesis	Tech. Disc. Form.	Book	Event Memories	Chap. book	Industrial prototype		Patent
Antioquia	Medellín, Abejorral, Amagá, Angelópolis, Angostura, Bello, Caicedo, Caramanta, El Carmen de Viboral, Cocomá, Donmatías, Entrerrios, Guarné, Itagüí, Jardín, Jericó, La Ceja, La Unión, Marimilla, Montebello, El Peñol, Retiro, Rionegro, Salgar, San Pedro de los Milagros, San Vicente Ferrer, Santa Bárbara, Sonsón, Tamesis, Urrao, Valparaiso, Venecia	43	74	21	9	5	10	6			168
Tolima	Alvarado, Anzoátegui, Espinal, Falan, Fresno, El Guamo, Herveo, San Sebastián de Mariquita, Planadas	14	15	13	4	6	4			56	
Valle del Cauca	Cali, Caicedonia, Candelaria, Cartago, Alcalá, Palmira, Restrepo, Sevilla, Trujillo, Tuluá, Ulloa	16	11	10	2	1	2			42	
Risaralda	Pereira, Apía, Belén de Umbria, Guática, La Celia, Marsella, Santa Rosa de Cabal, Santuario	13	9	4	2	1	1	1		31	
Cundinamarca	Anolaima, Cachipay, Sasaima, Silvanía, Supatá, Tibacuy, Tibirita, Tocaima, Villeta, Viotá, Soacha, Chía, El Rosal, Mosquera, Fusagasugá, Caparrapi, El Colegio, Funza, Guadua, Pandí, Pasca, San Francisco, Bogotá D.C	16	3	2		1	1	3		26	
Cauca	Corinto, Morales, Píndamo - Tunía, Sotará - Paispamba, Timbio, Toribio, El Tambo, Poyán, Argelia, Cajibío	10	9	4		2	1			26	
Caldas	Manizales, Neira, Marquetalia, Chinchiná, Belalcázar, Aranzazu, Anserma, Norcasia, Villamaría, Victoria, Samaná, Palestina, Pácora	3	13	2	2	2	2	1		25	
Bolívar	Córdoba, El Carmen de Bolívar, María la Baja, San Jacinto, San Juan Nepomuceno, Zambrano	8	6	2			1	1		18	
Santander	Lebríja, Piedecuesta, Barichara, Capitanejo, Cerrito, San Vicente de Chucurí, Socorro, Tona, Valle de San José	11	6	1			1			18	
Quindío	Armenia, Circasia, Finlandia, Pijao, Quimbaya, Salento	6	5	3	1	1				16	
Huila	Isnos, La Argentina, Palestina, Pitalito, Salado-blanco, Timaná, Acevedo	6	3			1	1	1		12	
Sucre	Sincelejo, Colosó, Chalán, Ovejas, Palmito, San Onofre, San José de Tolúviejo, Morroa		7	1			1	1		10	

Continue

Continuation	Dept.	Municipalities	Number of documents										TOTAL		
			Undergraduate work	Scientific article	Thesis	Tech. Disc. Form.	Book	Event Memories	Chap. book	Industrial prototype	Patent				
	Norte de Santander	Bochalema, Cáchira, Cucutilla, El Zulia, Labateca, Toledo	6	2											8
	Meta	Villavicencio, Acacias, Granada, Puerto López	4	1	1										6
	Cesar	Chiriguana, Curumani, La Jagua de Ibirico, Paillitas, San Alberto, San Diego, Valledupar, Agustín Codazzi, La Paz, Pueblo Bello, San Martín	1	4											5
	Nariño	Sandoná, San Bernardo, Pasto, Albán, Taminango, Chachagüí, El Tablón de Gómez, Leiva, Iles	1	3				1							5
	La Guajira	La Jagua del Pilar, Urumita, Riohacha, Dibulla, San Juan del Cesar, Manaure	1	1	1										3
	Magdalena	No informado		1	1										2
	Boyacá	San Pablo de Borbur, Garagoa	2												2
	Chocó	Quibdó		1											1
	Córdoba	San Carlos	1												1
	Amazonas	Uninformed		1											1
	Uninformed	Review article	83	43	15	14	11	4	9	6	2				187
	General total		245	224	80	35	31	25	27	6	2				675

Source: Authors

Note: there are searches that were carried out in more than one department.

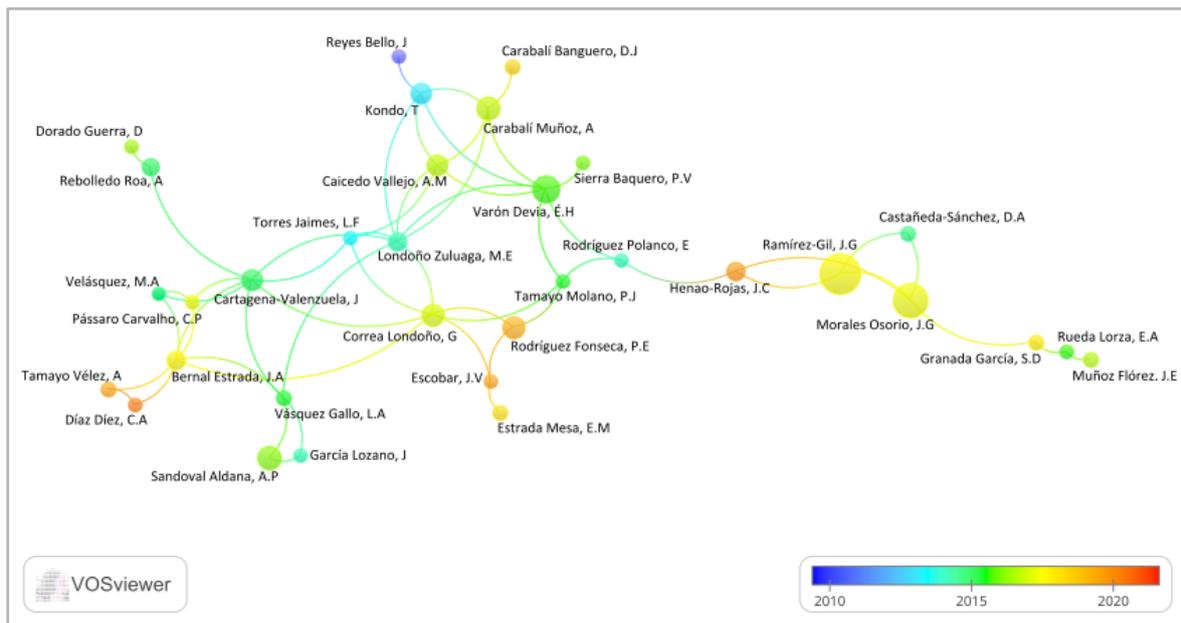


Figure 4: Bibliometric network of co-authorship linked to avocado in Colombia (14.2% density and 12.7% degree of centrality). Search date: April 2021.

highest degree of centrality were Hass (23), mercadeo (22) and transformación (22). Cluster 1 presented the least current information; however, it had the largest number of documents, whose central theme was the transformation of the fruit. The information in group 2 was the most current and the marketing theme prevailed. However, the most relevant word was Hass (Table 3). The other three clusters presented a low number of linked terms; however, they are predominant for the subsector. The transformation of the fruit was the third topic with the highest centrality and also the one with the highest number of links in the clusters analyzed.

Web search trends

Google Trends showed an upward trend in avocado (Figure 5.A). The term “aguacate” has been more consistent in the search for images with several peaks in the search window (Figure 5.B). The words with the highest relevance in the trend search in Google Trends were ‘ensalada’, ‘pelo’, ‘Hass’ and ‘aceite’. When the information was disaggregated by content type, we found that the behavior varies with related words and department (Table 4).

The highest value in audiovisual content on YouTube was recorded in November 2020: At this time, a trend was presented due to a video linking avocado with a high potassium content (Figure 5.C). The search for avocado in

Google Shopping reached its first popularity peak on 13 July 2019, when the first negotiations for avocado export to Asia were held (Figure 5.D).

The search in the news showed that the term was highly popular in specific periods, varying due to commercial situations or the opening of markets (Figure 5.E). The results showed three dates of high popularity, February 2012, due to phytosanitary problems in avocado crops caused by the winter wave associated with the ENSO-La Niña and the controversy over the importation of avocado and other fruits. May 2019, the export of fresh and processed fruit was the main topic, with CorpoHass exports to Kuwait, the opening of the Santa Marta port for Hass avocado exports to the United States, and the issue of environmental sustainability in avocado crops in the Quindío department. Finally, in March 2021, with 90% popularity, due to the opening of the market to South Korea.

Finally, Twitter opinion trends responded to the market opening mainly for exports, or the growth of the subsector. In 2021, tweets with environmental concerns were identified due to references in Chile and Mexico about the improper management of water resources that created an environmental problem. In Colombia, the environmental issue also links Hass’s variety with protected areas, an aspect that is being regulated by regional autonomous corporations (Figure 5.F).

Table 3: Bibliometric clusters related to avocado in Colombia by keywords (46.6% density and 38.2% degree of centrality)

Cluster	Keywords
1	Avocado oil, vegetable oil, oils and fats, dyes, extraction, production, avocado pulp, seed and transformation
2	Agribusiness, marketing, foreign trade, exports, business plan
3	Food, quality, ripening, postharvest, sustainability
4	Education, entrepreneurship, internationalization, regional, Information and Communication Technologies (ICT's)
5	Biological control, Hass, disease management, pests

Search date: April 2021.

The distribution of the search for the term “aguacate” in the country was more popular in the Caldas department, where the topics related to the search for avocado are agricultural and cosmetic topics. The departments of Guainía, Amazonas, Vichada and Vaupés did not show results. Although Antioquia Department did not present the highest search volume in Google trend, it presented the most constant activity, seeing that since 2010 it has had an average popularity value.

DISCUSSION

The configuration of research networks and platforms for indexing results enables correction of duplication in research (Olsen *et al.*, 2017). A limitation found was the absence of repositories and databases. This reflects the need to start from a broad and updated state of the art to prevent the researcher from developing already existing knowledge (Flórez *et al.*, 2012). Horta & Santos (2020) found that in the field of natural and agricultural sciences, collaborative work is key in the construction of research agendas. Therefore, we suggest working in collaborative research networks to optimise resources in the generation of knowledge in the avocado sub-sector. Knowledge derived

from collaborative networks contributes to the promotion of competitiveness, social and cultural aspects, and the political development of rural organizations (Muñoz, 2013). This coincides with Cáceres-Zambrano *et al.* (2022), who suggest the generation of networks between science, academia, and guilds.

It is necessary to strengthen publication in the scientific sphere as well as to strengthen and disseminate information (Hernández Socha, 2020). Research in this subsector was found to be broad and diverse and requires a higher level of impact in the scientific community. Numerous metrics exist to assess the impact of publications (González-Pereira *et al.*, 2010; Guo *et al.*, 2021; Moed, 2010). The H-index is a frequent and generalised metric that contrasts the number of publications with the number of citations. (Hirsch, 2005). In this research, 68.2% of the articles identified are below the H-index of 100 and 24.6% are above this value. This is reflected in the improvement of knowledge quality and training processes in this field. In addition, higher-qualified professionals are made available for the subsector (Basso & di Tollo, 2022). It makes possible the innovation and optimization of processes and the use of resources (Djoumessi, 2021).

Table 4: Results from Google Trends and Twitter. Search window 2010 – 2021

	Words	Department
General	Salad, mask, hair care	Guaviare, Caldas, Arauca, Quindío, Huila
Images	Salad, drawings, tree, crop, fruit	Caquetá, Arauca, Casanare, Caldas, Risaralda
YouTube	Songs, hair, salad, potassium	Nariño, Putumayo, Risaralda, Huila, Cauca
Shopping	Fruit	La Guajira, Cauca, Risaralda, Cundinamarca, Antioquia
News	Fruit, plant, Hass	Casanare, La Guajira, Sucre, Tolima, Quindío
Tweets	Crop, water, Hass, export, Chile	Antioquia, Bogotá, Caldas, Santander, Bolívar, Valle del Cauca

Source: Authors

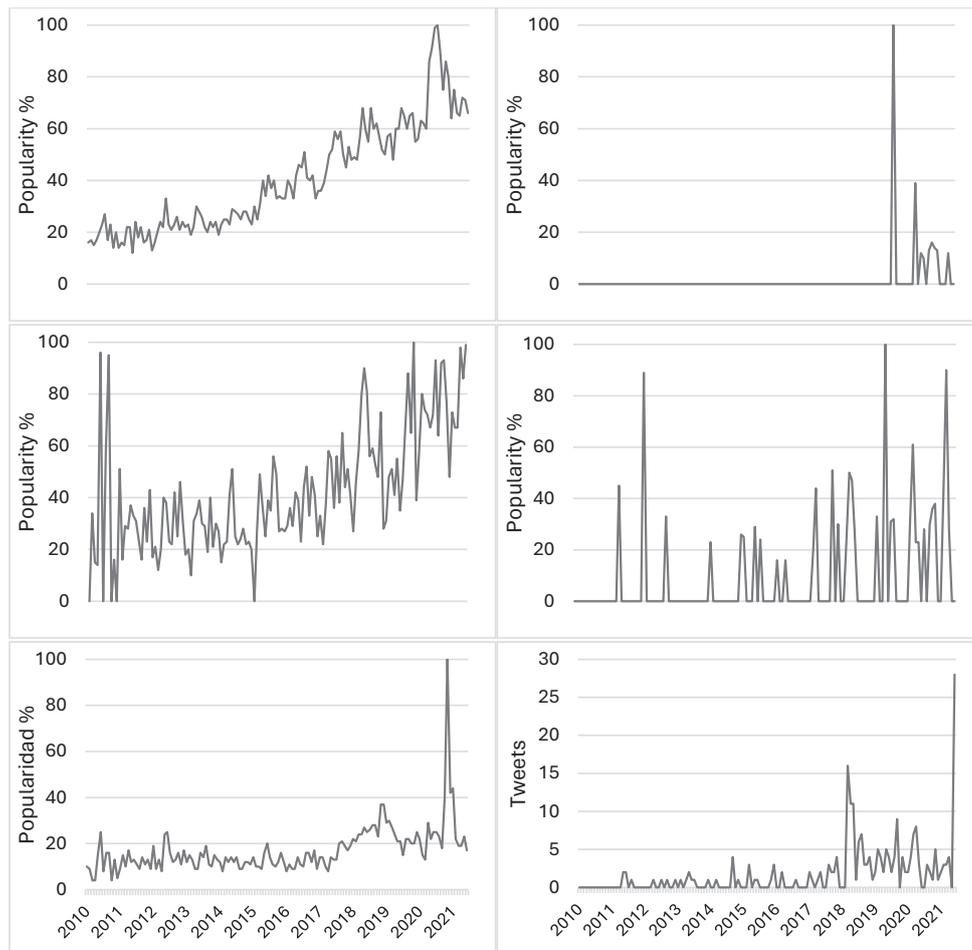


Figure 5: Web trends for the word “aguacate”, 2010 - 2021. A. General web search. B. Image search. C. Search on YouTube. D. Google shopping. E. Search on the News. F. Related Tweets. Source: Google Trends and Twitter.

Several authors highlight the importance of strengthening research in other areas with potential for the sub-sector, either for the Hass variety, or for criollos, green or papelillos such as the Montes de María subregion or other departments like Cesar, Santander and Norte de Santander (González *et al.*, 2018; Vásquez-Barajas & Bastos-Osorio, 2018). The data collected show the highest concentration of bibliographic information in the department of Antioquia, which is related to the growth in the avocado planted area (Ramírez-Gil, 2018). A relationship was found between the departmental distribution of the documents and the suitability map for the Hass avocado crop in Colombia (Fonseca *et al.*, 2018). However, the search carried out looked at the avocado fruit in general, so the geographical distribution of the results has a broader coverage, including areas with low altitudes where other varieties of avocados are grown, such as green, papelillos, and criollos. In the world, avocado research is concentrated in countries with the highest import volume (Cáceres-Zambrano *et al.*, 2022).

Locally, the lack of information in some eastern departments is possibly attributed to connectivity difficulties in connectivity (DANE, 2018; Ramírez & Blanco, 2020). Furthermore, it is a region with a significant proportion of protected areas and for this reason there may be a lack of interest in the crop (Jiménez *et al.*, 2019). The difficulties of accessing information and communication technologies in the country generate bias with respect to the interest raised in trend analysis. The need to collect regional information is highlighted, given that the requirements and opportunities vary depending on local characteristics.

However, we found that both the bibliometric analysis and the search trends are linked to the cosmetic industry, hair care, and dermatology. This has been verified empirically and experimentally (Cervantes-Paz & Yahia, 2021; Del Castillo-Llamosas *et al.*, 2021). Research with by-products helps reduce crop losses, process optimization and waste reduction (Dávila *et al.*, 2017; Hennessey-Ramos *et al.*, 2019).

Aguilar-Gallegos *et al.* (2022) found strong and weak links in the process of disseminating information and knowledge through social networks, which has traditionally been addressed in person. In addition, recognizing what the opinion trends in social networks allow to direct efforts towards topics of general interest. The perception in social networks of the value chain of avocados is related to the high-water consumption of the avocado crop, which is consistent with Caro *et al.* (2021). Environmental issues and water resources should be the cornerstones of the sector's development and ensure the sustainability of the system and the reduction of negative impacts in plantation areas. It should be coupled with the recognition of agricultural borders to avoid conflicts with conservation areas (Fonseca *et al.*, 2018).

The main challenge of technology and innovation in the avocado subsector is to link environmental aspects, food safety, sustainability, and trade. Participatory innovation processes are considered to be an opportunity to link various actors (Šūmane *et al.*, 2018). Innovation processes must allow the generation, access, exchange, and application of knowledge where different actors learn, innovate, order risks, and share benefits (Sonnino & Ruane, 2013).

Nowadays, access to information is increasingly simple, in addition, current dynamics show trends toward the digitization of knowledge, for the conservation, promoting access, opportunities for dissemination and improvement in information management (Soltis *et al.*, 2011; Wambari, 2017). Using the Web as a source of information is a research and market strategy in various sectors, even greater after the confinement situation due to the Covid-19 pandemic (Muhammad *et al.*, 2021; Rakshit *et al.*, 2021). Trend analysis is a methodology used to verify market needs and design attractive and commercial products (Wisetsri *et al.*, 2021). In this study, it was possible to identify the source of information that the network means for the academy and for science. The connection between offers and demands is often limited in the reception of needs and the transfer of knowledge. Using the Internet as a tool to overcome these gaps implies a challenge in connectivity and an opportunity to improve.

These types of methodology create the opportunity to broaden the research vision by using alternative networks for trend analysis and access to large amounts of data (Klerkx *et al.*, 2019). In this study, it was possible to find the usefulness of these methodologies for the agricultural sector in general. This analysis serves as a decision-making

tool and as a competitive strategy for the subsector.

This research found that the technological and research offer of the avocado subsector partially responds to the perception toward the value chain found in the trend analysis. Therefore, various investigations were found around marketing, exports, and environment, with little investigation of the last ones falling behind in this area. Likewise, our work shows that the use of data science tools allows access to a large amount of available information and that the correct filtering and analysis of data can provide more elements for decision making based on evidence. Furthermore, these tools constitute a key factor to understand the perception of citizens in a specific value chain, as well as to use them to improve this perception, and to gain knowledge of trends for the proper design of policies, products, and innovation.

Furthermore, our research is expected to contribute to evidence-based decision-making for key players in the avocado industry. Policymakers can leverage the findings to design targeted initiatives that promote technological advances and stimulate sustainable growth. Industry stakeholders can refine their strategies by aligning them with the prevailing public perceptions, ensuring that innovations resonate with consumer needs and expectations. Furthermore, academic researchers can identify novel research avenues that directly address the practical challenges faced by the avocado industry, thus accelerating its technological development. Our study embarks on a pioneering investigation of technological trends and perceptions in the avocado industry through the integration of bibliometric analysis and social media exploration.

CONCLUSIONS

The search carried out and the data analytics tools used in this study identified trends in research and development of the avocado subsector, as well as perception towards this production system. The search trends showed relevant topics for the subsector, including entering new markets, opportunities for product diversification, business development, and optimization of research processes through collaborative networks. It was found that the offers and demands are related; however, some demands are still unfulfilled. On the one hand, search trends (demands) are related to aspects of transformation, international trade, and the environment. The research (offers) partially responds to this, since the environmental area has been little explored. Then again, the type of document found has a still rudimentary level of development and dissemination, requiring

strengthening in the scientific sphere and technology transfer. Collaboration networks among authors were found to have the potential to be scaled to other actors in the chain, linking actors in charge of offers and demand. Using data science tools, the agricultural sector can learn about search dynamics and needs as a source of information in sectoral and business planning. As a final point, the actors in the value chain must contribute to the permanent updating of offers and demands, including different regions and themes.

ACKNOWLEDGEMENTS, FINANCIAL SUPPORT AND FULL DISCLOSURE

We thank the Universidad Nacional de Colombia for providing the academic and research environment that gave rise to this work. The authors have no conflict of interest to declare.

REFERENCES

- Aguilar-Gallegos N, Klerkx L, Romero-García LE, Martínez-González EG & Aguilar-Ávila J (2022) Social network analysis of spreading and exchanging information on Twitter: the case of an agricultural research and education centre in Mexico. *The Journal of Agricultural Education and Extension*, 28:115-136.
- Araújo RG, Rodríguez-Jasso RM, Ruiz HA, Pintado MME & Aguilar CN (2018) Avocado by-products: Nutritional and functional properties. *Trends in Food Science & Technology*, 80:51-60.
- Arias Bustos C & Moors EHM (2018) Reducing post-harvest food losses through innovative collaboration: Insights from the Colombian and Mexican avocado supply chains. *Journal of Cleaner Production*, 199:1020-1034.
- Aweke CS, Hassen JY, Wordofa MG, Moges DK, Endris GS & Rorisa DT (2021) Impact assessment of agricultural technologies on household food consumption and dietary diversity in eastern Ethiopia. *Journal of Agriculture and Food Research*, 4:100141.
- Basso A & di Tollo G (2022) Prediction of UK research excellence framework assessment by the departmental h-index. *European Journal of Operational Research*, 296:1036-1049.
- Borgatti SP, Everett MG & Freeman LC (2002) UCINET 6 for Windows: Software for social network analysis. Version 6.102, Analytic Technologies. Available at: <<https://ucinet.software.informer.com/6.0/>>. Accessed on: January 12th, 2021.
- Cáceres-Zambrano J, Jiménez-Hernández CN & Barrios D (2022) Tendencias en investigación y desarrollo tecnológico en la cadena productiva de aguacate (*Persea americana* L.). *Revista EIA*, 19:01-18.
- Callon M, Courtial JP & Laville F (1991) Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, 22:155-205.
- Caro D, Alessandrini A, Sporchia F & Borghesi S (2021) Global virtual water trade of avocado. *Journal of Cleaner Production*, 285:124917.
- Carranza DM, Varas-Belemmi K, de Veer D, Iglesias-Müller C, Coral-Santacruz D, Méndez FA, Torres-Lagos E, Squeo FA & Gaymer CF (2020) Socio-environmental conflicts: An underestimated threat to biodiversity conservation in Chile. *Environmental Science and Policy*, 110:46-59.
- Cervantes-Paz B & Yahia EM (2021) Avocado oil: Production and market demand, bioactive components, implications in health, and tendencies and potential uses. *Comprehensive Reviews in Food Science and Food Safety*, 20:4120-4158.
- DANE - Departamento Administrativo Nacional de Estadística (2018) *Indicadores Básicos de Tenencia y uso de Tecnologías de la Información y Comunicación*. Bogotá, DANE. 47p. (Boletín Técnico, 2022).
- Dávila JA, Rosenberg M, Castro E & Cardona CA (2017) A model biorefinery for avocado (*Persea americana* mill.) processing. *Bioresource Technology*, 243:17-29.
- Del Castillo-Llamosas A, del Río PG, Pérez-Pérez A, Yáñez R, Garrote G & Gullón B (2021) Recent advances to recover value-added compounds from avocado by-products following a biorefinery approach. *Current Opinion in Green and Sustainable Chemistry*, 28:100433.
- Dervis H (2019) Bibliometric Analysis using Bibliometrix an R Package. *Journal of Scientometric Research*, 8:156-160.
- Djournessi YF (2021) What innovations impact agricultural productivity in Sub-Saharan Africa? *Journal of Agriculture and Food Research*, 6:100228.
- Donthu N, Kumar S, Mukherjee D, Pandey N & Lim WM (2021) How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133:285-296.
- Fan W & Gordon M (2014) The power of social media analytics. *Communications of the ACM*, 57:74-81.
- Flórez DH, Morales A, Uribe CP & Contreras CA (2012) Análisis de tendencias en investigación básica para cadenas productivas agroindustriales. *Ciencia y Tecnología Agropecuaria*, 13:121.
- Fonseca F, Aguilar D, Siachoque R, Urbina J, Otero J, Páramo G, García E, Castillo Y, García E, Escobar C, Gómez J, Rodríguez L, Ayala J, Porras A, Enciso F, Latorre C, Ayala J, Camacho A, Acosta C, Osmos J & Lozano J (2018) Zonificación de aptitud para el cultivo comercial de aguacate Hass en Colombia, a escala 1:100.000. Unidad de Planificación Rural Agropecuaria. Available at: <https://upra.gov.co/es-co/Publicaciones/Zonas_aptas_aguacatehass.pdf>. Accessed on: February 23rd, 2021.
- FAO - Food and Agriculture Organization of the United Nations (2021) FAOSTAT Statistical Database. Available at: <<http://www.fao.org/faostat/en/#data/QC>>. Accessed on: March 8th, 2021.
- González GL, Figueroa Sánchez CI, Perdomo Ortiz MF & Ardila Rojas XA (2018) Producción de aguacate Hass una alternativa para el departamento del Huila. Eje temático: riesgo de mercado. *Revista Crecer Empresarial: Journal of Management and Development*, 02:07-15.
- González-Pereira B, Guerrero-Bote VP & Moya-Anegón F (2010) A new approach to the metric of journals scientific prestige: The SJR indicator. *Journal of Informetrics*, 4:379-391.
- Google (2021a) Ayuda de Tendencias de Búsqueda. Available at: <<https://support.google.com/trends/?hl=ko#topic=6248052>>. Accessed on: July 15th, 2021.
- Google (2021b) Ayuda de Google. Available at: <<https://support.google.com/?hl=es>>. Accessed on: July 15th, 2021.
- Guo X, Li X & Yu Y (2021) Publication delay adjusted impact factor: The effect of publication delay of articles on journal impact factor. *Journal of Informetrics*, 15:101100.
- Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C & Oliphant TE (2020) Array programming with NumPy. *Nature*, 585:357-362.
- He W, Shen J, Tian X, Li Y, Akula V, Yan G & Tao R (2015) Gaining competitive intelligence from social media data: Evidence from two largest retail chains in the world. *Industrial Management & Data Systems*, 115:1622-1636.
- Hennessey-Ramos L, Murillo-Arango W & Tovar Guayabo G (2019) Evaluation of a colorant and oil extracted from avocado waste as functional components of a liquid soap formulation. *Revista Facultad Nacional de Agronomía Medellín*, 72:8855-8862.
- Hernández Socha Y (2020) Scientific encounters between Colombia and the United States analyzed through publishing practices in *Caldasia* journal: The birds of the Republic of Colombia as a publishing event.

- Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 82:101289.
- Hirsch J (2005) An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102:16569-16572.
- Honnibal M & Montani I (2017) spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. Available at: <<https://spacy.io/usage/v2/>>. Accessed on: August 12th, 2021.
- Horta H & Santos JM (2020) The Multidimensional Research Agendas Inventory - Revised (MDRAI-R): Factors shaping researchers' research agendas in all fields of knowledge. *Quantitative Science Studies*, 1:60-93.
- ITC - International Trade Centre (2021) Trade statistics for international business development, Trademap. Available at: <<https://www.trademap.org/Index.aspx>>. Accessed on: March 8th, 2021.
- Jiménez CJG, Mantilla CLM & Barrera GJA (2019) Enfoque Agroambiental: Una mirada distinta a las intervenciones productivas en la Amazonia, Caquetá y Guaviare. *Instituto SINCHI*, 5:44.
- Jun SP, Yoo HS & Choi S (2018) Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, 130:69-87.
- JustAnotherArchivist (2020) Snsrape: A social networking service scraper in Python. Available at: <<https://github.com/JustAnotherArchivist/snsrape>>. Accessed on: August 12th, 2021.
- Khan HU, Nasir S, Nasim K, Shabbir D & Mahmood A (2021) Twitter trends: A ranking algorithm analysis on real time data. *Expert Systems with Applications*, 164:113990.
- Klerkx L, Aarts N & Leeuwis C (2010) Adaptive management in agricultural innovation systems: The interactions between innovation networks and their environment. *Agricultural Systems*, 103:390-400.
- Klerkx L, Jakku E & Labarthe P (2019) A review of social science on digital agriculture, smart farming and agriculture 4.0: New contributions and future research agenda. *NJAS: Wageningen Journal of Life Sciences*, 90-91:100315.
- Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, Kelley K, Hamrick J, Grout J, Corlay S, Ivanov P, Avila D, Abdalla S, Willing C & Jupyter development team (2016) Jupyter Notebooks – a publishing format for reproducible computational workflows. In: Loizides F & Schmidt B (Eds.) *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. Amsterdam, IOS Press. p.87-90.
- Kumar A, Kumar Mangla S & Kumar P (2022) An integrated literature review on sustainable food supply chains: Exploring research themes and future directions. *Science of The Total Environment*, 821:153411.
- McKinney W (2010) Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*, 1:56-61.
- Mendeley (2020) Guides - Explore Mendeley tools, community and discover the huge range of groups, people and institutions that use our tools. Available at: <<https://www.mendeley.com/guides>>. Accessed on: March 8th, 2021.
- Ministerio de Educación Nacional (2020) Instituciones de educación superior. Available at: <<https://www.datos.gov.co/Educaci-n/MEN-INSTITUCIONES-EDUCACI-N-SUPERIOR/n5yy-8nav>>. Accessed on: February 24th, 2021.
- Moed HF (2010) Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, 4:265-277.
- Moral-Muñoz JA, Herrera-Viedma E, Santisteban-Espejo A & Cobo MJ (2020) Software tools for conducting bibliometric analysis in science: An up-to-date review. *Multidisciplinar*, 29:e290103.
- Muhammad SS, Dey BL, Kamal MM & Syed Alwi SF (2021) Consumer engagement with social media platforms: A study of the influence of attitudinal components on cutting edge technology adaptation behaviour. *Computers in Human Behavior*, 121:106802.
- Muñoz M (2013) Transferencia tecnológica para el sector rural; la responsabilidad de un reto que genera oportunidades. *Desarrollo & Gestión*, 9:31-36.
- Olsen L, Kelly J & Kopriva N (2017) The agriculture network information collaborative (Agnic): Building on the past, looking to the future. *Library Trends*, 65:279-292.
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P & Moher D (2021) The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 29:372.
- Tirado-Kulieva VA, Gutiérrez-Valverde KS, Villegas-Yarlequé M, Camacho-Orbegoso EW & Villegas-Aguilar GF (2022) Research trends on mango by-products: a literature review with bibliometric analysis. *Journal of Food Measurement and Characterization*, 16:2760-2771.
- Quintero Ramirez S, Marín Sánchez B, Cubillos Jimenez S, Ruiz Castañeda W & Giraldo Ramirez D (2019) Avocado and Coffee Supply Chains Specialization in Colombia. *Procedia Computer Science*, 158:573-581.
- Rakshit S, Mondal S, Islam N, Jasimuddin S & Zhang Z (2021) Social media and the new product development during COVID-19: An integrated model for SMEs. *Technological Forecasting and Social Change*, 170:120869.
- Ramírez JM & Blanco C (2020) El internet como un aspecto prestacional del derecho a la educación. Bogotá, IDEP. 74p.
- Ramírez-Gil JG (2018) Avocado wilt complex disease, implications and management in Colombia. *Revista Facultad Nacional de Agronomía Medellín*, 71:8525-8541.
- Ramírez-Gil JG, Franco G & Henao-Rojas JC (2019) Review of the concept of quality in Hass avocado and the pre-harvest factors that determine it under tropical conditions. *Revista Colombiana de Ciencias Hortícolas*, 13:359-370.
- Ramírez-Gil JG & Morales-Osorio JG (2021) Diseases and disorders associated with different stages of crop development and factors that determine the incidence in Hass avocado crops. *Revista Ceres*, 68:71-82.
- Soltis P, Nelson G & James S (2011) Green digitization: Online botanical collections data answering real-world questions. *Applications in Plant Sciences*, 6:e1028.
- Sonnino A & Ruane J (2013) La innovación en agricultura como herramienta de la política de seguridad alimentaria: el caso de las biotecnologías agrícolas. In: de Jaramillo EH (Ed.) *Biotecnologías e innovación: el compromiso social de la ciencia*. Colombia, Pontificia Universidad Javeriana. p.25-52.
- Šūmane S, Kunda I, Knickel K, Strauss A, Tisenkopfs T, Rios I, Rivera M, Chebach T & Ashkenazy A (2018) Local and farmers' knowledge matters! How integrating informal and formal knowledge enhances sustainable and resilient agriculture. *Journal of Rural Studies*, 59:232-241.
- Tomlins S (2020) Scraping Tweets by Location in Python using snsrape. Available at: <<https://medium.com/swlh/how-to-scrape-tweets-by-location-in-python-using-snsrape-8c870fa6ec25>>. Accessed on: August 12th, 2021.
- Van der Aalst W (2016) *Data Science in Action*. In: van der Aalst W (Ed.) *Process Mining*. Berlin, Springer. p.03-23.
- Van Eck NJ & Waltman L (2014) *Visualizing Bibliometric Networks*. In: Ding Y, Rousseau R & Wolfram D (Eds.) *Measuring Scholarly Impact: Methods and Practice*. Berlin, Springer. p.285-320.
- Vásquez-Barajas EF & Bastos-Osorio LM (2018) Análisis de metodologías para la evaluación de cadenas de valor: caso del aguacate en Norte de Santander. *Respuestas Journal of Engineering Sciences*, 23:66-69.
- Vinholis MDMB, Saes MSM, Carrer MJ & de Souza Filho HM (2021) The effect of meso-institutions on adoption of sustainable agricultural technology: A case study of the Brazilian Low Carbon Agriculture Plan. *Journal of Cleaner Production*, 280:124334.

- Wambari C (2017) Digitization and Preservation of Academic Library Collections in Kenya: Issues, Opportunities and Lessons from Developing Societies. Available at: <<https://repository.cuk.ac.ke/xmlui/bitstream/handle/123456789/218/Wambari-KLISC2017-PDF.pdf?sequence=1&isAllowed=y>>. Accessed on: February 24th, 2021.
- Wisetsri W, Ragesh ST, Aarthy CJC, Thakur V, Pandey D & Gulati K (2021) Systematic Analysis and Future Research Directions in Artificial Intelligence for Marketing. *Turkish Journal of Computer and Mathematics Education*, 12:43-55.
- Wright E, Khanfar NM, Harrington C & Kizer LE (2010) The Lasting Effects of Social Media Trends on Advertising. *Journal of Business & Economics Research*, 14:75-82.
- Zakaria R, Vit P, Wijaya A, Ahmad AH, Othman Z & Mezzetti B (2022) A bibliometric review of *Persea americana* Mill. (Lauraceae): A green gold in agroindustry. *AIMS Agriculture and Food*, 7:831-854.