Universitat de Girona

THREE BIBLIOMETRIC ANALYSIS ARTICLES ON THE APPLICATION OF COMPOSITIONAL ANALYSIS THEORY (CoDa) IN THE SOCIAL SCIENCE

Carolina Navarro-López

Universitat de Girona

DOCTORAL THESIS

THREE BIBLIOMETRIC ANALYSIS ARTICLES ON THE APPLICATION OF COMPOSITIONAL ANALYSIS THEORY (CoDa) IN THE SOCIAL SCIENCE

Carolina Navarro López

2022

Programa de Doctorat en Dret, Economia i Empresa

Dirigida por: Carles Mulet Forteza i Salvador Linares Mustaròs

Tutor: Joan Carles Ferrer Comalat

Memòria presentada per optar al títol de doctor per la Universitat de Girona

Thesis as a compendium of publications

The three articles that constitute the present doctoral thesis are the following:

1) Navarro-López, C., González-Morcillo, S., Mulet-Forteza, C., & Linares-Mustarós, S. (2019). A bibliometric analysis of the 35th anniversary of the paper "The Statistical Analysis of Compositional Data" by John Aitchison (1982). *Austrian Journal of Statistics*, *50*, 38–55.

   The Austrian Journal of Statistics is an indexed journal in the Scopus database. In 2019, the journal had an impact factor of 0.589 (SJR) which ranked it in the second quartile of all those areas in which it was indexed, i.e. Applied Mathematics (position 222 out of 608 journals), Statistics, Probability and Uncertainty (position 71 out of 165 journals) and Statistics and Probability (position 177 out of 250 journals). The paper received one citation from a journal indexed in the Web of Science.

2) Navarro-López, C., Linares-Mustarós, S., & Mulet-Forteza, C. (2022). The statistical analysis of compositional data by John Aitchison (1986): A bibliometric overview. *SAGE Open*, *12*(2) 1–17. DOI: 10.1177/21582440221093366

   Sage Open is a journal indexed in the Social Sciences Citation Index database of the Web of Science. In 2021, the most recent year for which information is available, the journal had an impact factor of 2.032 (JCR), which placed it in the second quartile of the Social Sciences, Interdisciplinary area (position 53 out of 111 journals), the highest position ever achieved by the journal.

3) Navarro-López, C., Linares-Mustarós, S., & Mulet-Forteza, C. Research progress in compositional data in social sciences. A bibliometric analysis. Paper submitted to the journal Mathematical Geosciences.

   Mathematical Geosciences is a journal indexed in the Social Sciences Citation Index Expanded database of the Web of Science. In 2021, the most recent year for which information is available, the journal had an impact factor of 2.508 (JCR), which placed it in the second quartile of the Mathematics, Interdisciplinary Applications area (position 36 out of 108 journals).

Research area: Entrepreneurship, accounting and finance.

Els Doctors Salvador Linares Mustaròs i Carles Mulet Forteza

DECLAREM:

Que el treball titulat "THREE BIBLIOMETRIC ANALYSIS ARTICLES ON THE APPLICATION OF COMPOSITIONAL ANALYSIS THEORY (CoDa) IN THE SOCIAL SCIENCE", que presenta Carolina Navarro López per a l'obtenció del títol de doctor, ha estat realitzat sota la nostra direcció.

I, perquè així consti i tingui els efectes oportuns, signem aquest document.

Salvador Linares Mustaròs Carles Mulet Forteza

Girona, 5 de setembre de 2022.

Girona
Setembre 2022

Dedicat a Joel, al meu marit Roberto, als meus pares, a Carles i Salvador

Acknowledge

I want to thank my son, Joel, for all his support and the strength he has given me to fight for all my projects. To my husband, Roberto, for his immense patience and support. Without the two of them all this would not have made sense. And, thanks to my parents for always being there.

I also want to thank all the support and advice of the directors Salvador Linares and Carles Mulet. To Salvador for his follow up and professionalism as thesis director. To Carles, in particular, who has been involved beyond imagination and with whom I have been friends for 21 years, thanks for all the support in this project and in many others, a great professional and a better person. Without either of them this project could have been carried out and I would not have finished this doctoral thesis. Finally, I would like to extend my thanks to the Universitat de Girona and the Universitat de les Illes Balears.

INDEX

1. Resum

L’anàlisi de dades composicionals és una metodologia de recent aplicació amb grans possibilitats d’expansió a múltiples disciplines científiques.

La present tesi doctoral pretén ajudar a la difusió i a la transferència de coneixement sobre la importància adquirida en l’aplicació de la metodologia CoDa en les diferents categories de recerca, oferint informació sobre l’estat actual d’aquestes, els possibles col·laboradors i els temes candents. Conseqüentment, es proporcionarà orientació rellevant als investigadors que desitgin desenvolupar futures recerques en aquesta disciplina.

Amb tal motiu, s’ha dut a terme una anàlisi de diversos aspectes imprescindibles per a oferir un cos realista de l’avaluació i ús d’aquesta tècnica en les diferents categories científiques indexades en la Web of Science (WoS) des de 1982 fins a 2022.

D’aquesta manera, al llarg de les publicacions s’han anat destacant dades tan significatives com les universitats més productives, els autors més influents, ho temis que comencen a ser rellevants, entre altres aspectes, amb la finalitat d’oferir una informació més precisa a la comunitat científica. D’altra banda, i amb la finalitat de presentar nous coneixements relacionats tant amb les dades composicionals com amb la bibliometria, en la present tesi s’ha proposat un nou mètode d’anàlisi bibliomètrica a través de tècniques avaluatives relacionals per a proporcionar major claredat tant en les anàlisis realitzades com en els aspectes estudiats.

Les conclusions obtingudes en els tres articles estan associades a la investigació bibliomètrica sobre l’aplicació de la metodologia CoDa i basades en els documents publicats per John Aitchinson (1982, 1986), així com la seva aplicació específica en el camp de les ciències socials.

En el primer article, A bibliometric analysis of the 35th anniversary of the paper "The Statistical Analysis of Compositional Data de John Aitchinson (1982), es va realitzara una anàlisi bibliomètrica exhaustiva de l’esmentat document.

Pel que fa al segon article, "The statistical análisis of compositional data of John Aitchinson (1986): A bibliometric overview", es va realitzar una anàlisi bibliomètrica de totes les publicacions que han citat el llibre.

Finalment, en el tercer article, "Research Progress in compositional data in Social Science. A bibliometric analysis", es van analitzar els principals indicadors en l’àrea específica de les ciències socials.

D’aquesta manera, es pot observar com la recent aplicació de la metodologia basada en l’anàlisi composicional de dades ha està brindant l’oportunitat que els estudis abordin les necessitats i objectius més demandats i necessitats per la societat actual i per les institucions a nivell internacional.

Paraulas claus: John Aitchinson, anàlisi de dades compositives, bibliometria, Web of Science, classificació d'autors, classificació d'institucions, classificació de països, VOSviewer, estructura de cites, classificació de revistes, temes candents, anàlisi

bibliométrica, dades de composició (CoDa), resultats de la investigació, cartografia científica, Software VOSviewer.

2. **Resumen**

El análisis de datos composicionales es una metodología de reciente aplicación con grandes posibilidades de expansión en múltiples disciplinas científicas.

La presente tesis doctoral pretende ayudar a la difusión y a la transferencia de conocimiento sobre la importancia adquirida en la aplicación de la metodología CoDa en las diferentes categorías de investigación, ofreciendo información sobre el estado actual de las mismas, los posibles colaboradores y los temas candentes. Consecuentemente, se proporcionará orientación relevante a los investigadores que deseen desarrollar futuras investigaciones en esta disciplina.

Con tal motivo, se ha llevado a cabo un análisis de diversos aspectos imprescindibles para ofrecer un cuerpo realista de la evaluación y uso de esta técnica en las diferentes categorías científicas indexadas en la Web of Science (WoS) desde 1982 hasta 2022.

De esta forma, a lo largo de las publicaciones se han ido destacando datos tan significativos como las universidades más productivas, los autores más influyentes, lo temas que comienzan a ser relevantes, entre otros aspectos, con el fin de ofrecer una información más precisa a la comunidad científica. Por otra parte, y con el fin de presentar nuevos conocimientos relacionados tanto con los datos composicionales como con la bibliometría, en la presente tesis se ha propuesto un nuevo método de análisis bibliométrico a través de técnicas evaluativas relacionales para proporcionar mayor claridad tanto en los análisis realizados como en los aspectos estudiados.

Las conclusiones obtenidas en los tres artículos están asociadas a la investigación bibliométrica sobre la aplicación de la metodología CoDa y basadas en los documentos publicados por John Aitchinson (1982, 1986), así como su aplicación específica en el campo de las ciencias sociales.

En el primer artículo, A bibliometric analysis of the 35th anniversary of the paper “The Statistical Analysis of Compositional Data de John Aitchinson (1982), se realiza un análisis bibliométrico exhaustivo del citado documento.

En cuanto al segundo artículo, “The statistical análisis of compositional data of John Aitchinson (1986): A bibliometric overview”, se realizó un análisis bibliométrico en todas las publicaciones que han citado el libro.

Por último, en el tercer artículo, “Research Progress in compositional data in Social Sciene. A bibliometric analysis”, se analizó los principales indicadores en el área específica de las ciencias sociales.

De esta forma, se puede observar cómo la reciente aplicación de la metodología basada en el análisis composicional de datos está brindando la oportunidad de que los estudios aborden las necesidades y objetivos más demandados y necesitados por la sociedad actual y por las instituciones a nivel internacional.

Palabras clave: John Aitchinson, análisis de datos compositivos, bibliometría, Web of Science, clasificación de autores, clasificación de instituciones, clasificación de

países, VOSviewer, estructura de citas, clasificación de revistas, temas candentes, análisis bibliométrico, datos de composición (CoDa), resultados de la investigación, cartografía científica, Software VOSviewer.

3. Abstract

Compositional data analysis is a recently applied methodology with great potential for expansion in multiple scientific disciplines.

This doctoral thesis aims to help the dissemination and transfer of knowledge about the importance acquired in the application of CoDa methodology in the different research categories by providing information about the current state of the art, potential collaborators and hot topics. Consequently, relevant guidance will be provided to researchers wishing to develop future research in this discipline.

For this reason, an analysis has been carried out of several essential aspects in order to offer a realistic body of the evaluation and use of this technique in the different scientific categories indexed in the Web of Science (WoS) from 1982 to 2022.

In this way, throughout the publications, significant data such as the most productive universities, the most influential authors, the topics that are beginning to be relevant, among other aspects, have been highlighted in order to offer more precise information to the scientific community. On the other hand, and in order to present new knowledge related to both compositional data and bibliometrics, this thesis has proposed a new method of bibliometric analysis through relational evaluative techniques to provide greater clarity both in the analyses carried out and in the aspects studied.

The conclusions obtained in the three articles are associated with bibliometric research on the application of the Coda methodology and based on the papers published by John Aitchinson (1982, 1986), as well as its specific application in the field of social sciences.

In the first article, A bibliometric analysis of the 35th anniversary of the paper "The Statistical Analysis of Compositional Data by John Aitchinson (1982), an exhaustive bibliometric analysis of the aforementioned document was carried out. ¡

Regarding the second article, "The statistical analysis of compositional data of John Aitchinson (1986): A bibliometric overview", a bibliometric analysis was carried out in all the publications that have cited the book. ¡

Finally, the third article, "Research Progress in compositional data in Social Sciene. A bibliometric analysis", analyzed the main indicators in the specific area of the social sciences.

In this way, it can be observed how the recent application of the methodology based on compositional data analysis is providing the opportunity for studies to address the needs and objectives most demanded and needed by today's society and by institutions at the international level.

Keywords: John Aitchinson, compositional data analysis, bibliometrics, Web of Science, author ranking, institution ranking, country ranking, VOSviewer, citation structure, journal ranking, hot topics, bibliometric analysis, compositional data (CoDa), research output, science mapping, VOSviewer Software.

4. General Introduction

Synthesising the findings of research in a field of study advances knowledge in that field. Scholars have developed various scientific methods for this purpose, the most widely used being the quantitative approach of bibliometric research methods (Schmidt, 2008) and the scientific mapping of such research methods, which make it possible to determine the structure and development of scientific fields and disciplines.

Bibliometric analyses include the quantification of scholarly output on the basis of certain classifications that project indirect indications of its perception. Thus, bibliometrics quantifies, among many other aspects, the citation structure of an academic field. Moreover, and this is the most relevant aspect, bibliometrics allows mapping its own results by revealing the connections of the analysed field (Gumpenberger, Wieland & Gorraiz, 2012; Vogel, 2014).

Bibliometrics focuses exclusively on the measurement of publications. However, the term publication is relatively ambiguous as it can contain, among others, book chapters, journal articles and proceedings in conference volumes. Therefore, before starting a bibliometric investigation, it should be clearly defined what is being measured and what type of publication should serve as the basis for the bibliometric analyses being conducted.

Based on the above, bibliometrics should provide information about all the key components of research, among which we can highlight, just as an example, the analysis of the structure of scholarly activities in individual disciplines, academic productivity broken down to country level, the influence of countries or regions on particular knowledge domains, institutional collaboration, scientific output and its influence, etc. (Moed, De Bruin, Nederhof, Van Raan & Tijssen, 1992).

In this PhD thesis, various bibliometric analyses have been applied to compositional data analysis (CoDa) methods, which have their origins in geology and chemistry. Chemical analyses typically focus on the relative importance of the parts of a rock or substance being analysed. Chemical composition is therefore the focus of interest, while the sample size is largely irrelevant. Following the seminal works of Aitchison (1982, 1986), forty years of development have led to a well-established standard toolbox for compositional analysis in a multitude of areas, especially those related to the social sciences. Typical examples where CoDa methodology has been applied in the field of social sciences are, among others, geology, economics, medicine, food industry, ecology, palaeontology, agriculture, and sociology.

CoDa was born as a response to the problems encountered in applying standard statistical methods to part-whole data with frequently constant sum (Aitchison, 1986) leading to violation of key assumptions. However, it took a long time to find a solution to the problem of how to perform a proper statistical analysis of this type of data, i.e. to solve the spurious correlation problem, as Karl Pearson called it in 1897, or the closure problem, as Felix Chayes called it in the 1960s. Since standard statistical techniques lose their applicability and classical interpretation when applied to compositional data, new techniques were needed. No theoretically sound solution was proposed until the 1980s when John Aitchison proposed a coherent theory based on logarithms. From this, it is possible to rigorously develop any statistical analysis (cluster analysis, discriminant analysis, factor analysis and regression models, to mention just a few). Since the publication of John Aitchinson's article (1982) and his subsequent book in 1986, the application of this methodology has grown exponentially.

Recently, the emphasis on CoDa has shifted from the problem of the constant sum and statistical assumption violation to a strong interest in relative magnitudes. The term compositional analysis (Barceló-Vidal & Martín Fernández, 2016) has been coined to refer to the fact that it is the analysis of relative magnitudes, driven by the practitioner's goals and research questions, which make an analysis compositional, rather than the fact that the data may constitute parts of a whole or have a constant sum.

As noted above, the use of standard statistical techniques on component data can generate inconsistent results due to several undesirable problems, including problems of spurious correlation with proportions (Pearson, 1897), scale dependencies, presence of outliers and biases, out-of-range predictions (negative or greater than a constant sum) or inconsistent sub-combinations (Aitchison, 1986). In 1986, Aitchison published a book called The Statistical Analysis of Compositional Data (Aitchison, 1986), which detailed a whole set of techniques based on compositional data, and the results obtained were consistent as they were based on solid mathematical foundations. Today, these foundations remain valid and provide a solid basis for validating results, which is why Aitchison's (1986) book is universally considered to be both essential and seminal.

This is why it was necessary to quantify and analyse this growth and find out which authors, institutions and countries are most productive and efficient in this field of research, given the importance it has acquired and how relatively current this statistical application is. In this way and through bibliometrics, the aim is to quantify the research that has been carried out in the different categories reflected in the Web of Science using the compositional data analysis (CoDa) methodology.

In this doctoral thesis, a bibliometric analysis is presented through three publications, the first one in the first publication 'A bibliometric analysis of the 35th anniversary of the paper "The Statistical Analysis of Compositional Data" by John Aitchison (1982)', to commemorate the 35th anniversary of the publication of the seminal article. Subsequently, another analysis is conducted to realise an exhaustive bibliometric analysis of all publications that have cited the article based on data extracted from the Web of Sciences (WoS) regarding 'The Statistical Analysis of Compositional Data (1986)', to study the existing relationships between this seminal publication in the field of geoscience and modern science. Finally, after observing how compositional data analysis has expanded since the publication of Aitchinson's seminal work in 1982, it was decided to analyse the research area of the social sciences to gain a deeper understanding of the evolution that has occurred in this area.

The three articles that constitute the present doctoral thesis are the following:

1) Navarro-López, C., González-Morcillo, S., Mulet-Forteza, C., & Linares-Mustarós, S. (2019). A bibliometric analysis of the 35th anniversary of the paper "The Statistical Analysis of Compositional Data" by John Aitchison (1982). *Austrian Journal of Statistics*, *50*, 38–55.

The Austrian Journal of Statistics is an indexed journal in the Scopus database. In 2019, the journal had an impact factor of 0.589 (SJR) which ranked it in the second quartile of all those areas in which it was indexed, i.e. Applied Mathematics (position 222 out of 608 journals), Statistics, Probability and Uncertainty (position 71 out of 165 journals) and Statistics and Probability (position 177 out of 250 journals). The paper received one citation from a journal indexed in the Web of Science.

2) Navarro-López, C., Linares-Mustarós, S., & Mulet-Forteza, C. (2022). The statistical analysis of compositional data by John Aitchison (1986): A bibliometric overview. *SAGE Open*, *12*(2) 1–17. DOI: 10.1177/21582440221093366

Sage Open is a journal indexed in the Social Sciences Citation Index database of the Web of Science. In 2021, the most recent year for which information is available, the journal had an impact factor of 2.032 (JCR), which placed it in the second quartile of the Social Sciences, Interdisciplinary area (position 53 out of 111 journals), the highest position ever achieved by the journal.

3) Navarro-López, C., Linares-Mustarós, S., & Mulet-Forteza, C. Research progress in compositional data in social sciences. A bibliometric analysis. Paper submitted to the journal Mathematical Geosciences.

Mathematical Geosciences is a journal indexed in the Social Sciences Citation Index Expanded database of the Web of Science. In 2021, the most recent year for which information is available, the journal had an impact factor of 2.508 (JCR), which placed it in the second quartile of the Mathematics, Interdisciplinary Applications area (position 36 out of 108 journals).

The first two articles are the result of the joint research experience in bibliometrics and CoDa methodology of the co-directors of the thesis, Salvador Linares-Mustarós and Carles Mulet-Forteza. As a result of their knowledge, the need arose to carry out a study on the evolution and importance of the two seminal publications by John Aitchison (1982 and 1984) in order to determine how both works influenced the structure and development of the scientific disciplines in CoDa. Finally, the third paper has gone a step further by investigating the evolution of the use and application of this CoDa methodology in the area of social sciences.

5. Methodology

The methodology of the three studies was common and was based on various bibliometric analyses. These were established based on a procedure carried out in several stages. Firstly, the research questions were defined to develop the work. Secondly, the most appropriate bibliometric database was selected. Subsequently, and thirdly, the criteria for the search for the information to be analysed were established. Once the information was obtained, and fourthly, the information obtained was adequately filtered to be able to apply the most appropriate bibliometric methods. Finally, by analysing the results obtained, the development of the CoDa research field was traced in order to advance knowledge and propose new lines of research.

The first decision to be made was the selection of databases. In our work, and despite the availability of other options, we decided to use the Web of Science (WoS), as it is considered to be the most influential database in the world (Merigó et al., 2015). WoS is currently the most widely used database for bibliometric analysis. It is a multidisciplinary database containing 65 million records from more than 12,500 journals, organised into approximately 230 categories. This database is updated weekly and provides extensive bibliographic information on the publications indexed in it, including abstracts of these publications (although only since 1991), citations obtained by these documents, author, references cited, institutions and countries involved in the document, the journal in which the document was published, the journal's impact index, etc. Its most famous index is the Science Citation Index (SCI), although in recent years it has also introduced an extended version of the index, the Science Citation Index Extended (SCIE), which brings together more than 9,000 journals in 178 categories, while its social science edition, the Social Science Citation Index (SSCI) brings together information from almost 3,400 journals in almost 60 categories. However, the reality is that only more than 5% of academic journals are indexed by WoS. The WoS editorial team propagates this as a standard of quality, which means that the vast majority of scholarly work worldwide is ignored.

In the following, we describe the main bibliometric analyses and methods that have been carried out. These methods introduce a systematic, transparent and reproducible review process of the analysed publications that should allow the application of both performance analysis (also called evaluative techniques) and scientific mapping (also called relational techniques) (Cobo, López-Herrera, Herrera-Viedma & Herrera, 2011; Benckendorff & Zehrer, 2013). Performance analysis (Hall, 2011) includes productivity metrics such as the number of cited articles, the number of articles per author, etc. On the other hand, scientific mapping or relational techniques attempt to determine the knowledge structure of scientific disciplines, as well as determine their conceptual structure (Boyack, Klavans & Börner, 2005; Zupic & Cater, 2015). The main relational techniques are co-citation analysis, keyword co-occurrence analysis, co-author analysis and bibliographic linking. These relational techniques help researchers to identify intellectual and social structures in scientific fields, the emergence of new research topics and methods, and the identification of co-citation and co-authorship patterns (Benckendorff & Zehrer, 2013; Leydesdorff & Vaughan, 2006; Nerur, Rasheed & Pandey, 2015; Pilkington & Lawton, 2014; Ronda-Pupo & Guerras-Martín, 2012; Tan & Ding, 2015). Despite all this, both evaluative and relational techniques are unable to directly measure the quality of academic research, but rather deduce it indirectly through its impact history.

In terms of bibliometric indicators in evaluative techniques, the following have been used in this doctoral thesis:

- Number of publications: this provides information on the academic performance of a researcher, institution, country, etc. It is an indicator of productivity but does not indicate influence. This indicator forms the basis for the other bibliometric indicators.

- Number of citations received by a publication: this attempts to measure the relevance of the publication. However, this assumes that there is some correlation between the number of citations by a paper and its quality. This is not always true, however, as sometimes a publication may be cited to criticise it or to indicate that it is poorly produced. Bibliometric studies also often provide an analysis of the number of citations obtained in a given field of research, usually in the form of lists of a given number of the most cited authors, papers or journals in the field under analysis. Thus, taking into account all the above considerations, the number of citations reveals the impact of a publication in its field of research. For this reason, citation thresholds are often set, e.g. more than 10, 100 or 1,000 citations, to make it easier to infer the quality of the publication. One of the main limitations of this bibliometric indicator is produced by self-citation, i.e. those citations that also come from researchers who cite their document, thus increasing the number of citations, which can sometimes be detrimental to the prestige of the indicator.

- Number of citations per document: this is obtained by dividing the total number of citations by the total number of publications made by a researcher, an institution, a country, etc. In this respect, however, it should be borne in mind that researchers often cite documents they have produced themselves, the so-called self-citations. Therefore, in the analysis of citations per document, or even total citations, the number of self-citations is usually eliminated to make comparisons more objective.

- H-Index: in 2005, Jorge E. Hirsch developed an indicator that, like the number of total citations per document, combines the number of publications and their citation frequency in a single index.

Based on the above metrics, the following bibliometric analyses were carried out:

a) Analysing the structure of publications and citations in a scientific field.
b) Analysis of the most productive and most cited authors.
c) Analysis of the most productive and most cited institutions.
d) Analysis of the most productive countries and the most cited.

After applying assessment techniques, it is recommended to generate a bibliometric network map. Scientific mapping produces a spatial representation and visualisation of findings regarding the structure and dynamics of scientific fields, so it can be considered a combination of classification and visualisation (Boyack & Klavans, 2014). The analyses performed included co-citation, co-occurrence, co-authorship and bibliographic linking:

- Co-citation analysis uses citation counts to construct measures of similarity between documents, authors or journals (McCain, 1990). According to Small (1973), co-citation refers to the frequency with which two documents are cited together in the citation reference of other documents, so if at least a third document cites two documents, these documents are said to be co-cited. There are different types of co-citation, including co-citation analysis of authors, papers and journals containing the co-cited papers (McCain, 1990; White & Griffith, 1981; White & McCain, 1998).

- Bibliographic linking occurs when two papers contain at least one common reference to a third paper in their bibliographies (Kessler, 1963). Bibliographic coupling is used as a measure of similarity between documents (Zupic & Cater, 2015), which becomes greater as the number of common references between documents increases. Bibliographic linking is useful in a wide variety of fields as it helps researchers to find related past research, as well as to discover the directions of an academic discipline, and to demonstrate the possible barriers and boundaries of research (Ma, 2012).

- Co-occurrence of keywords is understood as the joint occurrence of two terms in a specific text. This technique analyses the content of a text on the basis of the joint occurrence of pairs of words that make it possible to identify relationships between concepts within a given domain. The greater the frequency of joint occurrence of the words, the greater their conceptual linkage (Miguel, Caprile & Jorquera-Vidal, 2008). In this respect, it is the only relational technique that uses the actual content of documents to construct a measure of similarity, since the other techniques connect documents indirectly through citations or co-authorships. Keyword analysis can be applied to document titles, keywords, abstracts and even to the full texts of analysed documents.

- Co-authorship analysis analyses the levels of scientific collaboration (Acedo, Barroso, Casanueva & Galan, 2006) since it examines the networks created when researchers collaborate in the elaboration of scientific papers (Acedo, Barroso, Casanueva & Galan, 2006), making it of special interest for analysing research topics involving scientific collaboration (Benckendorff, 2009; Racherla & Hu, 2010; Ye Li & Law, 2013). A co-authorship relationship is established when two or more authors jointly produce a paper (Lu & Wolfram, 2012).

With the application of the above bibliometric techniques, the doctoral thesis shows a network map of the related indicators. This allows a more convenient tracing of the initial theoretical roots and historical context of the field. Furthermore, the keyword analysis approach adopted through temporal evolution allows researchers to track the development of CoDa in the literature, providing opportunities to expand the current body of research (Liu et al., 2022), enabling the generation of new approaches as well as the identification of future research trends. In summary, we identified the main research fields and topics cited by Aitchison's two papers (1982, 1986), in order to subsequently apply these bibliometric methods to the field of social sciences.

6. General objectives and objectives of each article

The aim of bibliometrics is to measure the output side of science. By adding scientific mapping or relational techniques, it is possible to determine the intellectual structure of a scientific discipline, as well as to establish and determine the conceptual structures of these disciplines (Boyack, Klavans & Börner, 2005; Zupic & Cater, 2015). Thus, the main objective of this doctoral thesis is based on the identification of the patterns of production and efficiency of bibliographic information by scientists from different institutions worldwide in the application of the compositional data analysis (CoDa) methodology in any field of research related to the social sciences. The first objective is to carry out an analysis of the references and works published by scientists in order to determine the current relevance of the bibliography they use, the journals they consult and the research fronts that may be associated with these journals. The second objective is to establish a representation of the network structure of the application of the CoDa methodology to the different areas of research in those publications that have cited the works of John Aitchinson (1982, 1986), in order to subsequently make it more specific to the social sciences area. In order to achieve these aims, three different research projects were carried out, where the main objective of the present doctoral thesis was proposed and covered.

Concerning the specific objectives of each of the works carried out, these are summarised below:

1) The specific objectives of the first article, **A bibliometric analysis of the 35th anniversary of the paper "The Statistical Analysis of Compositional Data" by John Aitchison (1982)**, were:

    a) What is the evolution of the number of citations of Aitchison's 1982 paper?
    b) Who are the authors who cite Aitchison's work the most?
    c) Which institutions cite Aitchison's work the most?

2) The specific aims of the second article, **The statistical analysis of compositional data by John Aitchison (1986): A bibliometric overview,** were:

    a) What is the scholarly structure of Aitchison's 1986 book?
    b) In which major journals has Aitchison's book been cited?
    c) What are the main themes discussed in the main papers published by authors citing Aitchison's book?

3) Finally, the specific objectives of the third article, **Research progress in compositional data in social sciences. A bibliometric analysis**, were:

    a) What is the academic structure of the application of CoDa methodology in the social sciences field?
    b) Which are the most productive authors in applying CoDa methodology in the social sciences?
    c) Which institutions are the most productive in publishing articles using CoDa methodology in the social sciences?
    d) Which countries have published the most papers using compositional data analysis methodology applied in the social sciences?

The results of our work clearly show the significant impact that John Aitchison's (1982, 1986) publications have had on scientific research, having been cited by authors and institutions publishing worldwide. As for the third paper, the results show the scientific categories that have published the most using compositional data analysis (CoDa), as well as the main collaborations between the most productive and influential authors, institutions and countries.

In conclusion, we can say that the objectives set have resulted in research that is useful for potential authors to have a quick snapshot of what is expected of events in this field of research.

7. Article #1

Austrian Journal of Statistics
January 2021, Volume 50, 38–55.
http://www.ajs.or.at/
doi:10.17713/ajs.v50i2.1066

A Bibliometric Analysis of the 35th anniversary of the paper "The Statistical Analysis of Compositional Data" by John Aitchison (1982)

Carolina Navarro-Lopez
University of the Balearic Islands

Silvia Gonzalez-Morcillo
University of the Balearic Islands

Carles Mulet-Forteza
University of the Balearic Islands

Salvador Linares-Mustarós
University of Girona

Abstract

This study presents a comprehensive bibliometric analysis of the paper published by John Aitchison in the Journal of the Royal Statistical Society. Series B (Methodological) in 1982. Having recently reached the milestone of 35 years since its publication, this pioneering paper was the first to illustrate the use of the methodology "Compositional Data Analysis" or "CoDA". By October 2019, this paper had received over 780 citations, making it the most widely cited and influential article among those using said methodology. The bibliometric approach used in this study encompasses a wide range of techniques, including a specific analysis of the main authors and institutions to have cited Aitchison' paper. The VOSviewer software was also used for the purpose of developing network maps for said publication. Specifically, the techniques used were co-citations and bibliographic coupling. The results clearly show the significant impact the paper has had on scientific research, having been cited by authors and institutions that publish all around the world.

Keywords: John Aitchison, compositional data analysis, bibliometrics, Web of Science, author ranking, institution ranking, country ranking, VOSviewer.

1. Introduction

Nowadays compositional data are defined as arrays of strictly positive numbers for which ratios between them are considered to be relevant (Egozcue and Pawlowsky-Glahn (2019)). Despite warnings about the problems involved in not using specific methods for such data (Pearson (1897), Chayes (1948) and Vistelius and Sarmanov (1961)), it was not until the 1980s that the first general methods were proposed as appropriate methods for their analysis (Aitchison (1982) and Aitchison (1986)). This methodology received the name of compositional data analysis, CoDa analysis or simply CoDA. It is usually written CoDa when it refers to "compositional data" and CoDA when it refers to "Compositional Data Analysis". That same terminology also encompassed methods that allow the analysis of data wich positive values, whereby although the data do not have to fulfill the characteristic of constant sum, they do need to meet the requirement that the study of certain ratios of this study is considered as relevant in the study of the problem. Sample applications of data that do not represent parts of any whole can be found in Rodrigues, Daunís-I-Estadella, Mateu-Figueras, and Thió-Henestrosa (2011), Ortells, Egozcue, Ortego, and Garola (2015) and Linares-Mustarós, Coenders, and Vives-Mestres (2018).

The scientific production related to compositional data analysis has increased constantly over the years, and in the last ten years especially started to flourish in very different fields to the ones where it was initially employed (Kogovšek, Coenders, and Hlebec (2013), Ferrer-Rosell, Coenders, and Martínez-Garcia (2015), Batista-Foguet, Ferrer-Rosell, Serlavós, Coenders, and Boyatzis (2015), Belles-Sampera, Guillen, and Santolino (2016), Morais, Thomas-Agnan, and Simioni (2018), Blasco-Duatis, Coenders, Saez, García, and Cunha (2019), Creixans-Tenas, Coenders, and Arimany-Serrat (2019), Carreras Simó and Coenders (2020) and Coenders and Ferrer-Rosell (2020)). This growth and expansion to new fields can be related to four easily identifiable events. Normally, after beginning its wanderings in an intuitive manner, every new scientific theory enters a period of mathematic axiomatic formalization. Set theory is a clear example of this. In this sense, the theory has not been an exception and the purely mathematic works in which the theory is axiomatized may have assigned a greater degree of confidence given by the scientific community towards the new methods (Pawlowsky-Glahn and Egozcue (2001), Martín-Fernández, Olea-Meneses, and Pawlowsky-Glahn (2001) and Egozcue, Pawlowsky-Glahn, and Gloor (2018)). The second event is the proliferation of tutorials and textbooks on the theory (Aitchison (1986), Aitchison (1994), Valls (2018), Pawlowsky-Glahn and Buccianti (2011), Van den Boogaart and Tolosana-Delgado (2013), Pawlowsky-Glahn, Egozcue, and Tolosana-Delgado (2015), Greenacre (2017) and Filzmoser, Hron, and Templ (2018)), which ensure its growth by facilitating the inclusion of new researchers. The third event that seems to have facilitated the expansion of CoDA consists in the development of multiple libraries with R (Van den Boogaart and Tolosana-Delgado (2013), Palarea-Albaladejo and Martín-Fernández (2015) and Filzmoser *et al.* (2018)) and the creation of various softwares, such as the CoDaPack (Thió-Henestrosa and Martín-Fernández (2005) and Comas-Cufí, Thió-Henestrosa, Egozcue, Tolosana-Delgado, and Ortego (2011)), which allows operations to be performed without any previous knowledge of programming. The fourth event likely to have triggered the expansion of the CoDa theory is the creation of different working groups, the multiple actions they have undertaken helping to disseminate this new theory. As examples of this, we can mention the establishing of a biannual congress, different introductory courses on CoDA and the creation of websites offering users, among other features, the option to download working material. Finally, it is also worth mentioning that the work done by the groups has led to the creation of a CoDa Association, spreading CoDa theory even wider.

To celebrate 35 years since publication of the seminal article on CoDa analysis "The Statistical Analysis of Compositional Data" (Aitchison (1982)), the main purpose of this paper is to carry out an exhaustive bibliometric analysis of all publications to have cited the paper based on data taken from the Web of Sciences (WoS). This analysis serves the purpose of gathering information on trends in research using CoDa analysis.

The bibliometric analysis is divided into two parts. The first presents an analysis of the academic structure used in the documents that have cited Aitchison's 1982 paper, while the second focuses on an analysis of the main authors, institutions and countries to have cited it. The information we expect to gather from the analysis should answer the following research questions (RQ):

- RQ1: What is the evolution of the number of citation of Aitchison's 1982 paper?
- RQ2: Who are the authors that most cite Aitchison's paper?
- RQ3: What are the institutions that most cite Aitchison's paper?
- RQ4: What are the countries that most cite Aitchison's paper?

To meet our aim, the WoS database and the VOSviewer Software (Van Eck and Waltman (2010)) were used. The VOSviewer Software was employed with the aim of graphically mapping the bibliographic material used. Specifically, the following techniques were considered in this paper: bibliographic coupling and co-citation. The reason for using the WoS database is that it is considered the most influential in the world (Merigó, Gil-Lafuente, and Yager (2015)).
The rest of the document is divided into the following sections: the second section presents the bibliometric methods used in this paper; the third section provides a complete bibliometric study of Aitchison's work "The Statistical Analysis of Compositional Data" (Aitchison (1982)); and the fourth section summarizes the main conclusions, limitations and future lines of research.

2. Methodology

The term bibliometrics was introduced by Pritchard (1969) as "the application of mathematical and statistical methods to books and other means of communication". Currently, although many other definitions exist (see Yuan, Gretzel, and Tseng (2015) and Köseoglu, Sehitoglu, Ross, and Parnell (2016)), they all describe it as an instrument for analyzing the evolution of scientific disciplines based on intellectual, social and conceptual structures (Zupic and Čater (2015)) in order to identify trends and patterns in scientific research (Merigó, Blanco-Mesa, Gil-Lafuente, and Yager (2017)). Therefore, bibliometrics is one of the most widely used approaches for analyzing how a scientific field develops (Bar-Ilan (2008)).
For this bibliometric study, data were gathered from the WoS database in October 2019 using "The Statistical Analysis of Compositional Data" as a keyword in the field "title" and "Aitchison, J." as a keyword in the field "author". These searches returned Aitchison's 1982 paper as the only result. Subsequently, the information was refined based on the total number of citations obtained by the paper, which resulted in 784 publications for analysis.
Given that no consensus exists in the literature on which methods are best or most appropriate, we used several bibliometric indicators to present the data. Firstly, we considered the number of publications and citations, these methods being considered the most popular according to Ding, Rousseau, and Wolfram (2016). The former indicates productivity, while the latter quantifies the influence of these publications (Svensson (2010)). Other common indicators include the most productive authors, institutions and countries, and number of publications and citations per person (Mulet-Forteza, Salvá, Monserrat, and Amores (2020)). For the analysis of institutions, we also included general university rankings. The results in the tables are sorted by total number of publications (TP).
In addition, we used the VOSviewer software (Van Eck and Waltman (2010)) to graphically map the bibliographic data (Sinkovics (2016)) for co-citations (Small (1973)) and bibliographic coupling (Kessler (1963)). Co-citation assumes that there is some kind of relationship between two documents cited jointly by a third document (McCain (1990), Ramos-Rodríguez and Ruíz-Navarro (2004) and Small (1973)). According to McCain (1986) and McCain (1991), these documents allow the academic structure of a scientific discipline to be determined. Bibliographic coupling measures the similarity of the subject analyzed among the documents considered from the frequency in which certain references are shared. A bibliographic coupling occurs when two documents include the reference to a third document (Young (1983)), so there is a possibility that these documents are linked (Martyn (1964)). Bibliographic coupling is usually applied to perform the graphic mapping of institutions and countries (Small (1999) and Boyack and Klavans (2014)), while co-citation is usually used to perform the graphic mapping of autors (Glänzel and Thijs (2012) and Zupic and Čater (2015)).
The combination of methods used to collect data from the WoS database, along with use of the VOSviewer software, allowed us to incorporate both the "full counting" and "fractional counting" techniques. The difference between these methods is that "full counting" assigns one point to each participant of a paper, whereas "fractional counting" takes into account co-

authorship of the paper (Mulet-Forteza, Genovart-Balaguer, Merigó, and Mauleon-Mendez (2019b)).

3. Bibliometric study of Aitchison's paper (1982)

In this section, we will address the different research questions posed.

3.1. Evolution of number of citations received by Aitchison's paper (1982)

Regarding the first question (RQ1), Figure 1 presents the evolution of the citations received by Aitchison's 1982 paper.

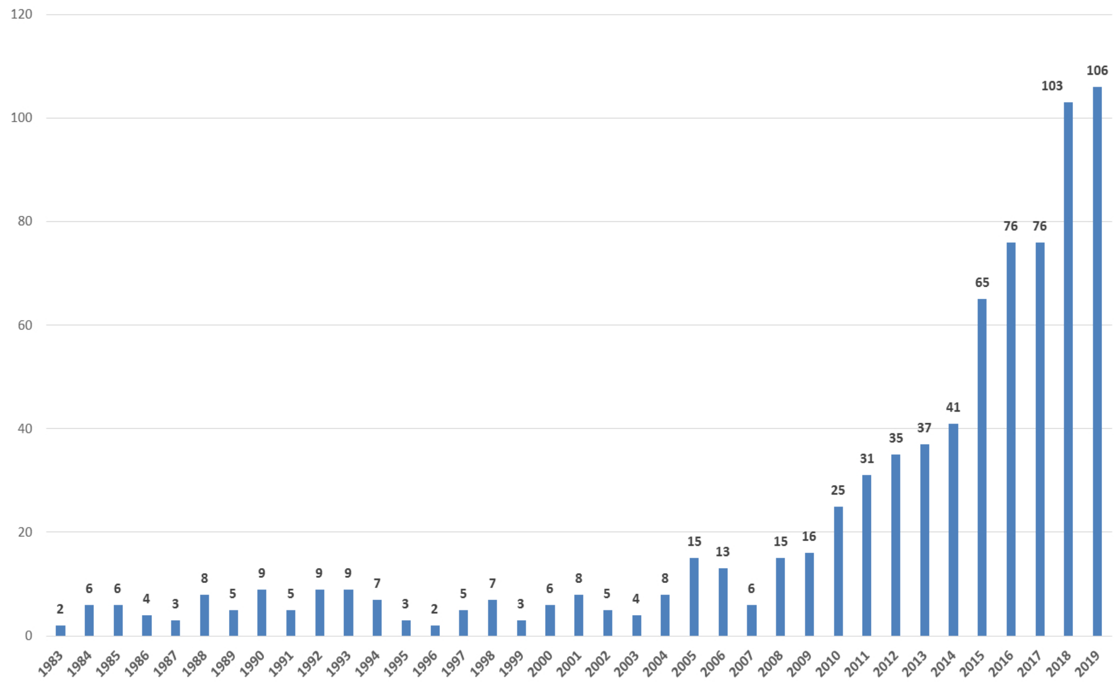

Figure 1: Annual number of citations received by Aitchison's 1982 paper. Source: own elaboration, compiled from WoS database.

Figure 1 shows that the paper has received uninterrupted citations since its publication in 1982. It also indicates how the number of citations received has evolved over different periods. In this sense, with few exceptions, the number of annual citations received by the paper between 1983 and 2007 did not exceed 10 per year. On the other hand, since 2008 annual citations have exceeded the previous value every year, following an expected exponential growth (Price (1986)). Likewise, a very significant increase in the number of citations received can be observed since 2015, and this increased still further in the years 2018 and 2019.
We have analyzed some of the reasons why Aitchison's 1982 paper has received a significant number of citations, especially since 2011. To this effect, in Table 1 we examine the evolution over time of the main research areas where the citations for the said work have been provided.

Table 1 clearly shows that the majority of citations received by Aitchison's 1982 paper come from three research areas, i.e. Mathematics, Geology and Environmental Sciences Ecology. Nevertheless, the interest of these research areas in Aitchison's 1982 paper has been aroused only in the last decade. Almost 60% of the citations from the area of Mathematics belong to this period, while the percentage goes up to over two-thirds in the areas of Geology and Environmental Sciences Ecology. Other areas that have also provided a great number of citations of Aitchison's 1982 paper are Geochemistry Geophysics, Mathematical Computational Biology, Engineering, Biochemistry Molecular Biology and Agriculture.

Table 1: Main research areas that have cited Aitchison's paper (1982). Source: own elaboration, WoS database, 1982 through December 2019. Abbreviations: R = Ranking that occupies the research area during the period 1983-2000; R1 = Ranking that occupies the research area during the period 2001-2009; R2 = Ranking that occupies the research area during the period 2010-2019; R3 = Ranking that occupies the research area during the whole period. Note: The same journal may be indexed in two or more research areas at the same time. Some examples are Mathematical Geology and Mathematical Geosciences, both indexed in the Geosciences and Mathematics research areas, and Bioinformatics, indexed in the Biotechnology and Mathematics research areas.

R1	R2	R3	R	Research Areas	1983-2000	2001-2009	2010-2019
1	1	1	1	Mathematics	55	36	130
2	2	2	2	Geology	27	25	103
3	3	3	3	Environmental Sciences Ecology	18	17	86
6	9	4	4	Computer Science	11	4	56
13	4	5	5	Geochemistry Geophysics	5	7	49
5	13	8	6	Mathematical Computational Biology	13	3	34
14	6	7	7	Engineering	5	6	36
8	33	6	8	Biochemistry Molecular Biology	9	0	37
15	7	10	9	Agriculture	5	6	27
4	5	16	10	Zoology	13	6	18
24	39	9	11	Science Technology Other Topics	4	0	33
23	18	11	12	Physical Geography	4	2	25
9	14	18	13	Chemistry	8	3	17
37	12	13	14	Water Resources	2	4	21
29	21	14	15	Public Environmental Occupational Health	3	1	19
78	78	12	16	Microbiology	0	0	23
38	19	17	17	Business Economics	2	2	18
17	15	20	18	Life Sciences Biomedicine Other Topics	5	3	12
51	57	15	19	Biotechnology Applied Microbiology	1	0	19
7	32	31	20	Nutrition Dietetics	11	0	7
22	11	33	22	Marine Freshwater Biology	4	4	7
16	10	40	24	Plant Sciences	5	4	5
48	16	26	25	Operations Research Management Science	1	3	10
19	20	28	26	Behavioral Sciences	5	1	8
47	8	36	28	Biodiversity Conservation	1	5	6
18	17	41	29	Paleontology	5	2	5
10	34	49	30	Endocrinology Metabolism	8	0	4
20	37	32	31	Physiology	5	0	7
11	35	89	42	Physical Sciences Other Topics	8	0	0
12	36	90	43	Reproductive Biology	7	0	0

On the other hand, Table 1 also illustrates a wide range of research areas that provided the largest amount of citations of Aitchison's 1982 paper when it was first published but are not so relevant today, including Zoology, Chemistry, Nutrition Dietetics, Plant Sciences, Behavioral Sciences, Paleontology, Endocrinology Metabolism, Physiology, Physical Sciences Other Topics, Reproductive Biology, among others.
It can therefore be seen that there has been a shift in interest in the research carried out by Aitchison in 1982, and that areas related to Statistics, Geosciences, Mathematics, Computer Science, Biochesmitry and Economics, among others, have replaced those initially used by CoDA. Consequently, the journals that have cited Aitchison's 1982 paper the most are those indexed in these research areas. Just as an example, it is noteworthy that the Journal of Geochemical Exploration, indexed in the Geochemistry Geophysics research area, is the one that has cited Aitchison's 1982 paper most often, with a total of 23 papers citing the said document during the last decade. Other journals indexed in the research areas that cited very often Aitchison's 1982 paper during the last 10 years include:

- In the Mathematics research area: Bioinformatics, Mathematical Geosciences, Environmental and Ecological Statistics and Environmetrics, among others.
- In the Environmental Sciences Ecology: Environmental Earth Sciences and International Journal of Environmental Research and Public Health, among others.
- In the Statistics research area: Journal of the American Statistical Association, Annals of Applied Statistics, Stochastic Environmental Research and Risk Assessment, Biometrics and Austrian Journal of Statistics, among others.
- In the Geosciences research area: Quaternary International and Geoderma, among others.

Therefore, it can be stated that the interest aroused in these research areas by Aitchison's 1982 paper has caused a genuinely growing interest in this publication, especially during the last decade.
It has also been possible to confirm that the authors who have most often cited Aitchison's 1982 paper during the last decade match those at the top of Table 2. In fact, only some positions have been exchanged. Thus, for example, Antonella Buccianti and Vera Pawlowsky-Glahn would exchange their positions, while Andrea Bloise, who occupies position 11 in Table 2, if we consider only the citations made to Aitchison's 1982 paper during the last decade, would occupy the ninth position in this new ranking, relegating John Aitchison from the TOP 10, who would be left out of the list of authors who have cited Aitchison's 1982 paper the most. This is not surprising, considering that John Aitchison died in 2016 at the age of 90.
Finally, we also analyzed the original source of the 784 citations received by Aitchison's paper. In this regard, 90.7% of citations were from documents published as papers, 4.6% from proceedings papers, 2.2% from books, 2% from reviews, and the remaining 0.5% from notes and letters. Thus, 93.2% of citations came from papers that had passed a strict process of arbitration; in other words, articles, reviews, letters and academic notes.

3.2. Most productive authors citing Aitchison's paper (1982)

In this section, we address the second question (RQ2) posed in our paper. Firstly, Table 2 lists those authors who have cited Aitchison's paper (1982) the most.

Vera Pawlowsky-Glahn (University of Girona, Spain) is the author who has cited the Aitchison paper (1982) the most, followed by Antonelle Buccianti (Università degli Studi di Firenze, Italy) and Juan José Egozcue (Polytechnic University of Catalonia, Spain). As Table 2 shows, the three main authors in this ranking have cited Aitchison's paper (1982) a total of 86 times, although it should be noted that this value, when obtained by means of a full counting

Table 2: Main authors who have cited Aitchison's paper (1982). Source: own elaboration, WoS database, 1982 through October 2019. Abbreviations: R = Ranking; TP = Total citing papers. Note: There are 25 authors with 4 papers.

R	Name	Institution (Country)	TP
1	Pawlowsky-Glahn, V	University of Girona (Spain)	32
2	Buccianti, A	Universiàt degli Studi di Firenze (Italy)	29
3	Egozcue, JJ	Polytechnic University of Catalonia (Spain)	25
4	Tolosana-Delgado, R	Helmholtz-Zentrum Dresden-Rossendorf (Germany)	18
5	Mateu-Figueras, G	University of Girona (Spain)	14
6	Aitchison, J	University of Glasgow (UK)	13
7	Martin-Fernandez, JA	University of Girona (Spain)	13
8	Li, HZ	University of Pennsylvania (USA)	11
9	Dumuid, D	University of South Australia (Australia)	9
10	Olds, T	University of South Australia (Australia)	9
11	Bloise, A	University of Calabria (Italy)	8
12	Gallo, M	Università degli Studi di Napoli L'Orientale (Italy)	8
13	Miriello, D	University of Calabria (Italy)	8
14	van den Boogaart, KG	Helmholtz-Zentrum Dresden-Rossendorf (Germany)	8
15	Woronow, A	Exxon Mobil Corporation (USA)	8
16	Palarea-Albaladejo, J	Biomathematics & Statistics Scotland (UK)	7
17	Weltje, GJ	KU Leuven (Belgium)	7
18	Crisci, GM	University of Calabria (Italy)	6
19	De Luca, R	University of Calabria (Italy)	6
20	Ortego, MI	Polytechnic University of Catalonia (Spain)	6
21	Wang, HW	Beijing Advanced Innovation Center for Big Data and Brain Computing (China)	6
22	Zuo, RG	China University of Geosciences, Wuhan (China)	6
23	Bagneres, AG	Université de Tours (France)	5
24	Blei, DM	Columbia University in the City of New York (USA)	5
25	Chaput, JP	University of Ottawa (Canada)	5
26	Gloor, GB	Western University (Canada)	5
27	Heslop, D	Australian National University (Australia)	5
28	Hron, K	Palacký University in Olomouc (Czech Republic)	5
29	Klotz, S	Helmholtz Zentrum für Umweltforschung (Germany)	5
30	Kuhn, I	Helmholtz Zentrum für Umweltforschung (Germany)	5
31	Lorenzi, MC	Universite Paris 13 (France)	5
32	Love, KM	BEB Erdgas und Erdförl GmbH (Germany)	5
33	Maher, C	University of South Australia (Australia)	5
34	Mueller, U	Edith Cowan University, Joondalup (Australia)	5
35	Parent, LE	Université Laval (Canada)	5
36	Scealy, JL	Australian National University (Australia)	5
37	Szava-Kovats, RC	Estonian Institute of Ecology (Estonia)	5
38	Tjallingii, R	Deutsches GeoForschungsZentrum (GFZ) (Germany)	5
39	Tremblay, MS	Children's Hospital of Eastern Ontario (Canada)	5
40	Tsagris, M	Panepistimio Kritis (Greece)	5
41	Wang, J	Chengdu University of Technology (China)	5
42	Welsh, AH	Australian National University (Australia)	5

method, does not take into account co-authors among these authors. This bias will be eliminated later when performing the graphic analysis of the main authors through a fractional counting method.

It is also interesting to observe the decreasing number of authors producing an increasing number of citing papers, as predicted by the bibliometric law of authors' productivity (Lotka (1926)).

The University of Calabria (Italy) counts four authors and is the most repeated institution among the authors who lead the ranking in Table 2, followed by the Australian National University, the University of Girona and the University of South Australia, with three authors each. With two authors, we find the Helmholtz Zentrum für Umweltforschung (Germany), the HZDR - Helmholtz-Zentrum Dresden-Rossendorf (Germany) and the Polytechnic University of Catalonia (Spain). The rest of the institutions have only one author represented in Table 2 (23).

Finally, the authors in Table 2 work in 13 different countries. Australia (with seven authors) leads this ranking, followed by Germany and Italy, with six authors each. Next, we find Spain (five authors), Canada (four authors), China and the US (with three authors each), France and the UK (with two authors each), while Belgium, Czech Republic, Estonia and Greece only have one author in Table 2.

Figure 2 shows a graphic map of the co-citations among the most influential authors to have cited Aitchison's paper.

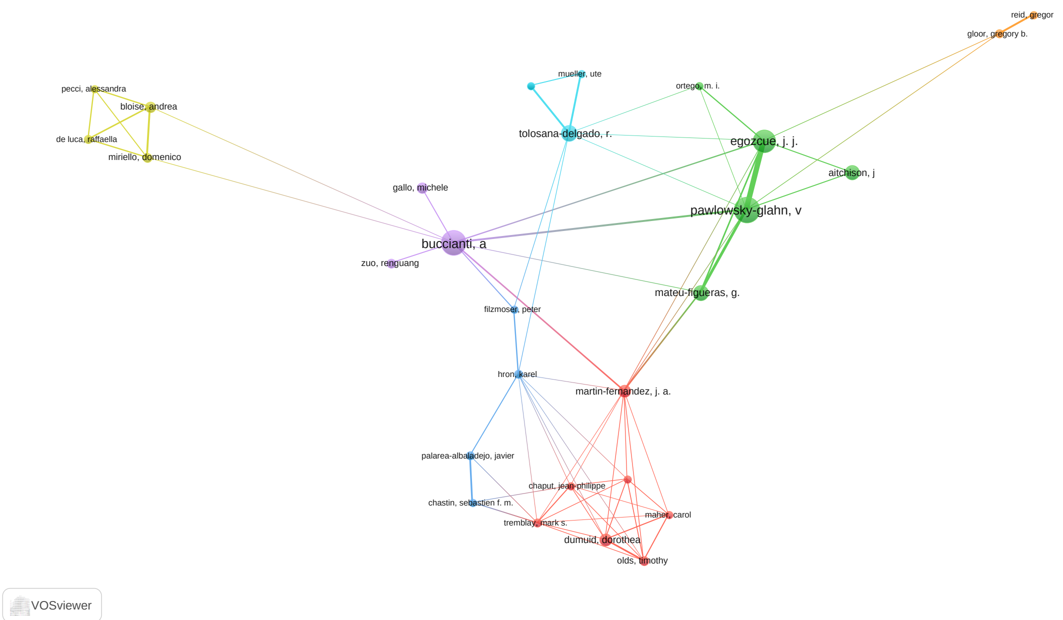

Figure 2: Co-citation of authors who have cited Aitchison's paper (1982). Node size = the number of citations received by an author; line thickness indicates multiple connections; line length is not significant. Citation threshold of 5 and showing the 100 most representative co-citation connections. Source: own elaboration, based on WoS database; figure created using VOSviewer Software.

Figure 2 reveals four main node clusters, indicating networks of connections between authors working on similar topics, and three secondary node clusters. The largest group, with seven authors, is focused around the figures of Dorothea Dumuid and Tim Olds, both from the University of South Australia. The second group contains five authors focused on the topics of Vera Pawlowsky-Glahn (University of Girona), Juan José Egozcue (Polytechnic University of Catalonia), Glòria Mateu-Figueras (University of Girona) and John Aitchison (University of Glasgow). This is the group with the largest network of connections, both between each other and with authors in the other nodes. The third group, consisting of four authors, is led by Javier Palarea-Albaladejo (Biomathematics & Statistics Scotland), while the fourth group, also with four authors, is led by Domenico Miriello and Andrea Bloise, both from the University of Calabria. With some exceptions, most authors in Figure 2 also appear in Table 2, which indicates that there are no significant differences between the analyses performed by the WoS database using the "full counting" method and that done by the VOSviewer Software using the "fractional counting" method.

3.3. Most productive institutions citing Aitchison's paper (1982)

Regarding the third question (RQ3) posed in our paper, Table 3 shows the institutions to have most frequently cited Aitchison's paper (1982), together with the position that these universities occupy in the Academic Ranking of World Universities (ARWU) (Consultancy (2019)) and the Quacquarelli Symonds World University Ranking (Symonds (2019)).

The University of Girona is the institution whose researchers have most frequently cited Aitchison's paper (1982), followed by the Polytechnic University of Catalonia and the University of Florence. The countries displaying the largest number of institutions in Table 3 are France (11), the UK (8) and the US (8). In addition, 12 institutions in Table 3 appear in the ARWU top 100, with Harvard University ranking the highest, in first position. Similarly, 13 universities appear in the top 100 in the QS ranking, with Harvard University leading the ranking, in third position on that list.
Table 3, which was compiled using the full counting method, can show biases in those institutions with cultures which encourage several authors to work together on a single paper. For this reason, Figure 3 shows the results of the previous analysis using the fractional counting method, eliminating the aforementioned bias.
Figure 3 shows a bibliographic coupling of the institutions that cite Aitchison's paper (1982).

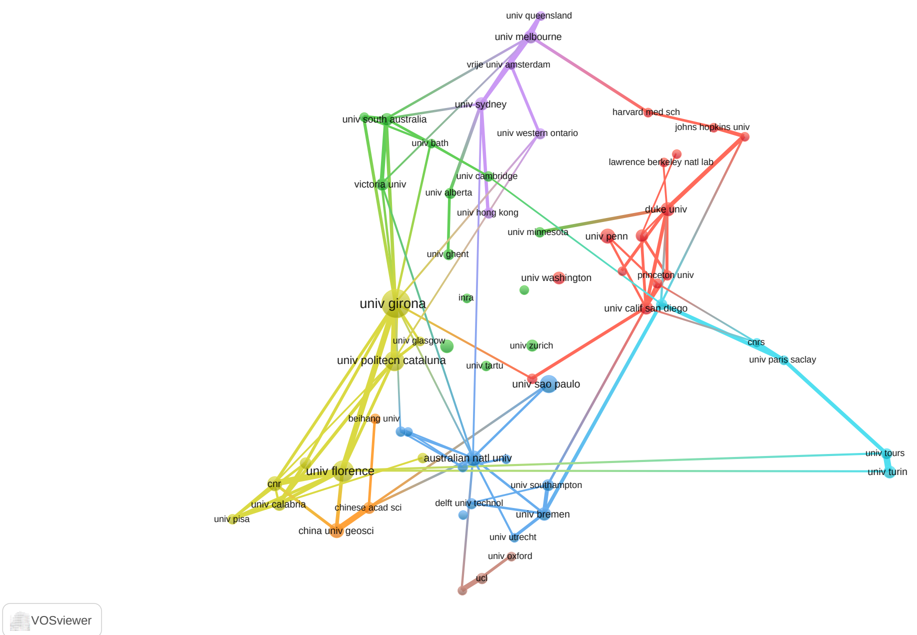

Figure 3: Bibliographic coupling of institutions citing Aitchison's paper (1982). Node size = number of citations received by authors belonging to a university; line thickness indicates multiple connections; line length is not significant. Citation threshold of five and showing the 100 most representative co-citation connections. Source: own elaboration, based on WoS database; figure created using VOSviewer Software.

Table 3: Main institutions that have cited Aitchison's paper (1982). Source: own elaboration, WoS database, 1982 through October 2019. Abbreviations: R = Ranking; TP = Total citing papers; QS = Quacquarelli Symonds (QS) World University Rankings (Symonds (2019)); ARWU = Academic Ranking of World Universities (Consultancy (2019)). Note: There are 21 institutions with seven papers.

R	Institution (Country)	TP	QS	ARWU
1	University of Girona (Spain)	46	-	801-900
2	Polytechnic University of Catalonia (Spain)	33	275	601-700
3	University of Florence (Italy)	31	501-510	201-300
4	Helmholtz Association (Germany)	29	-	-
5	Centre National de La Recherche Scientifique (France)	26	-	-
6	University of Sao Paulo (Brazil)	18	-	101-150
7	University of London, City (UK)	17	351	901-1000
8	Consiglio Nazionale Delle Ricerche CNR (France)	16	-	-
9	University of Hong Kong (China)	15	25	101-150
10	Australian National University (Australia)	14	24	76
11	China University of Geosciences (China)	14	-	401-500
12	Universite Confederale Leonard de Vinci (France)	14	-	-
13	Centre Val de Loire Comue (France)	13	-	-
14	Commonwealth Scientific Industrial Research Organisation (Australia)	13	-	-
15	Harvard University (USA)	13	3	1
16	James Hutton Institute (UK)	12	-	-
17	University of Pennsylvania (USA)	12	19	17
18	University of Sydney (Australia)	12	42	80
19	Duke University (UK)	11	26	28
20	Laval University (Canada)	11	402	201-300
21	University of Melbourne (Australia)	11	39	41
22	Communauté d'Universités et Établissements D'Aquitaine Comue (France)	10	-	-
23	United States Department of Energy Doe (USA)	10	-	-
24	University of Bremen (Germany)	10	501-520	501-600
25	University of Calabria (Italy)	10	-	801-900
26	University of California San Diego (USA)	10	41	18
27	University of South Australia (Australia)	10	264	501-600
28	CEA (France)	9	-	-
29	Institute for Humanities Social Sciences (France)	9	-	-
30	Institute of Ecology Environment (France)	9	-	-
31	Helmholtz Zentrum Dresden Rossendorf (Germany)	9	-	-
32	Institut National de La Recherche Agronomique (France)	9	-	-
33	University of Edinburgh (UK)	9	18	31
34	University of Southampton (UK)	9	96	101-150
35	University of Turin (Italy)	9	571-580	201-300
36	University of Washington (USA)	9	66	14
37	University of Washington Seattle (USA)	9	-	-
38	Istituto di Geoscienze E Georisorse (Italy)	8	-	-
39	Nerc Natural Environment Research Council (UK)	8	-	-
40	United States Department of the Interior (USA)	8	-	-
41	Universite Bourgogne Franche Comte Comue (France)	8	-	-
42	Universite Paris Saclay (France)	8	-	-
43	University College London (UK)	8	-	15
44	University of Barcelona (Spain)	8	166	151-200
45	University of Houston (USA)	8	651-700	210-300
46	University of Naples L'Orientale (Italy)	8	-	-
47	University of Oxford (UK)	8	5	7
48	University of Zurich (Switzerland)	8	78	61

Figure 3 shows four main node groups and five secondary node groups. The largest group of 14 institutions is focused around English-speaking institutions, including Harvard University, Duke University and the University of California San Diego. The second group contains 13 institutions, including the University of South Australia, Victoria University and the University of Zurich. The third is composed of 11 institutions, among which the University of Sao Paulo, the University of Turin and the University of Bremen stand out, and the fourth main node, with nine institutions, is led by the University of Girona, the Polytechnic University of Catalonia, and the University of Florence. The latter three lead the first three positions in Table 3. In turn, this group of institutions is the one with the largest network of connections, both with one another and with institutions in the other nodes. Therefore, with some exceptions, it can be observed that most of the institutions present in Figure 3 also occupy relevant positions in Table 3, which indicates that there are no significant differences between the analysis performed by the WoS database using the "full counting" method and the VOSviewer Software using "fractional counting".

3.4. Most productive countries citing Aitchison's paper (1982)

Regarding the fourth question (RQ4), Table 4 shows the countries that have most frequently cited Aitchison's 1982 paper.

Table 4 shows that the countries with the highest populations are not those that have cited Aitchison's 1982 paper most, with the exception of the US. As a matter of fact, only three countries in the top 10 (US, China and Brazil) have over 100 million inhabitants. In contrast, Table 4 shows how countries where English is widely spoken, especially among academics, are those where Aitchison's paper (1982) tends to be cited the most. This trend is especially present in countries such as Australia, Norway and New Zealand. These countries would lead the rankings in Table 4 if we ordered it based on number of papers cited by population.
Like the previous tables, Table 4 uses the full counting method, which skews countries where several authors write articles together rather than working independently. We therefore implemented a fractional counting method in Figure 4, which shows a bibliographic coupling of the countries that have cited Aitchison's 1982 paper most.

Figure 4 shows eight clusters. The first cluster by number of countries (10) is led by Germany and is composed of European countries, with the exception of Taiwan. The second cluster (seven countries) is led by Australia. This cluster is basically made up of non-European (6) countries. The third cluster by number of countries (6) is led by the United Kingdom. The United States leads the fourth cluster and is the most productive country, with the broadest network of connections on the map. The fifth cluster is led by Spain, while the last is led by Italy. In general, it can be observed that the results obtained under both the full counting system (WoS database) and the fractional counting system (VOSviewer Software) are very similar. Figure 4 shows a very diverse network of connections, where we find cultural connections between different countries such as Colombia and Ireland, or Australia and Iran.

4. Conclusions

Adopting a bibliometric approach and based on data obtained from the WoS database, in this paper we have carried out an analysis of all the publications that have cited the paper entitled "The Statistical Analysis of Compositional Data" published by John Aitchison in the Journal of the Royal Statistical Society. Series B (Methodological) in 1982. Having recently reached the milestone of 35 years since its publication, the paper is considered to be the seminal article on the CoDa analysis.
In this paper, we have met all of our established aims. Specifically, we have answered the four research questions we asked at the beginning. As for the first (RQ1), we have analyzed how the number of citations of this paper has evolved, showing how the paper has received uninterrupted citations since its publication and that over the past four years the number of

Table 4: Main countries that have cited Aitchison's paper (1982). Source: own elaboration, WoS database, 1982 through October 2019. Abbreviations: R = Ranking; TP = Total citing papers; POP = population in millions; TP/POP = total publications per millions of inhabitants Note: There are six countries with six papers.

R	Country	TP	POP	TP/POP
1	USA	207	327.2	0.63
2	UK	115	67.5	1.70
3	Spain	90	46.7	1.93
4	Italy	87	60.4	1.44
5	Australia	78	24.6	3.17
6	Germany	77	82.8	0.93
7	Canada	63	37.1	1.70
8	China	59	1403.4	0.04
9	France	53	67	0.79
10	Brazil	33	210.1	0.16
11	Netherlands	30	17.3	1.73
12	Belgium	21	11.4	1.84
13	Sweden	21	10	2.10
14	Switzerland	18	8.42	2.14
15	Norway	15	5.3	2.83
16	Czech Republic	13	10.6	1.23
17	New Zealand	13	4.9	2.65
18	Denmark	12	5.8	2.07
19	South Africa	12	56.7	0.21
20	Austria	10	8.8	1.14
21	Japan	9	126.8	0.07
22	Finland	8	5.5	1.45
23	Iran	8	82.6	0.10
24	Ireland	8	6.6	1.21
25	India	7	1372.1	0.01
26	Mexico	7	129.2	0.05
27	Russia	7	146.8	0.05
28	South Korea	7	51.5	0.14
29	Taiwan	7	23.8	0.29

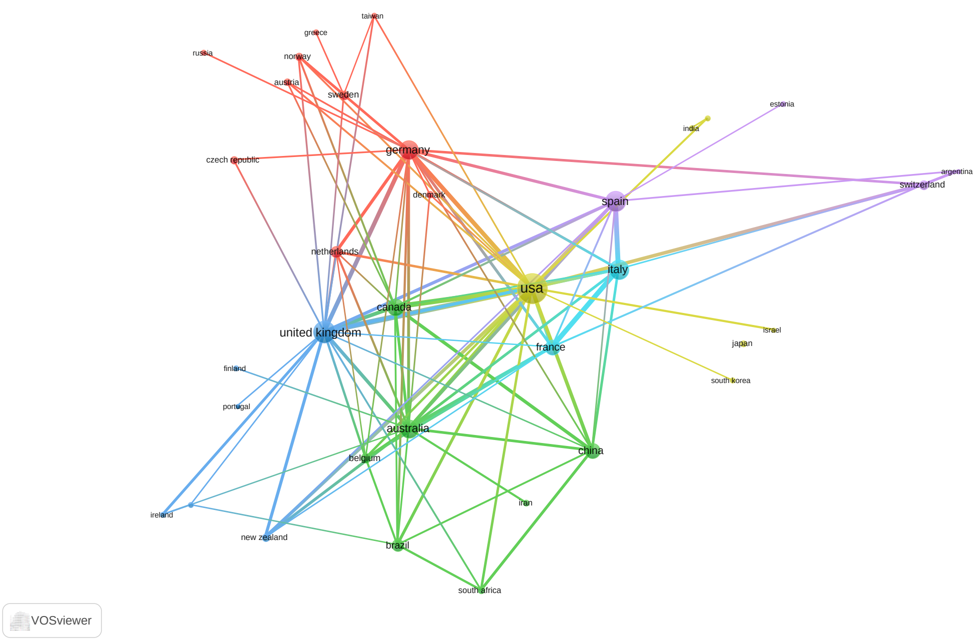

Figure 4: Bibliographic coupling of countries citing Aitchison's paper (1982). Node size = number of citations received by a country; line thickness indicates multiple connections; line length is not significant. Citation threshold of five and showing the 100 most representative co-citation connections. Source: own elaboration, based on WoS database; figure created using VOSviewer Software.

citations has increased more significantly in line with typical expected exponential growth. We have also corroborated that almost 95% of the 784 citations received by Aitchison's 1982 paper have come from documents that have undergone a strict arbitration process.

Regarding the second research question (RQ2), Vera Pawlowsky-Glahn and Glòria Mateu-Figueras (University of Girona), Antonella Buccianti (Università degli Studi di Firenze), Juan José Egozcue (Polytechnic University of Catalonia) and Raimon Tolosana-Delgado (HZDR - Helmholtz-Zentrum Dresden-Rossendorf) are the authors who have cited Aitchison's 1982 paper most. Furthermore, authors of the University of Girona, Polytechnic University of Catalonia, University of Florence, Helmholtz Association and Center National of La Recherche Scientifique CNRS are the ones who have cited the paper (RQ3) the most, while by country, the authors from the United States, the United Kingdom, Spain, Italy, Australia and Germany are the ones who have cited the paper (RQ4) the most. Our analysis indicates that there are no significant differences between the analysis of the WoS database using the "full counting" method and the "fractional counting" method used with the VOSviewer Software.

Although this document provides a description of the structure of citations, leading authors, institutions and countries that have cited Aitchison's 1982 paper, it does have some limitations. For example, since data were collected from the WoS database, the limitations of this database also apply to this analysis. As we have indicated previously, the WoS database collects information under a "full counting" method, meaning that documents with many co-authors generally have more weight than documents produced by a single author (Mulet-Forteza, Genovart-Balaguer, Mauleon-Mendez, and Merigó (2019a)). To resolve this limitation, we also employed the "fractional counting" method, using the VOSviewer software to identify co-citations and bibliographic coupling. A further limitation is that the results are dynamic and will inevitably change over time.

Despite the above limitations, this paper represents a starting point for future bibliometric studies in this field. In this respect, future lines of research should aim to carry out a bibliometric analysis focusing on all publications that have included the methodology of

"Compositional Data Analysis", firstly in the field of the social sciences, and then increasing the number of publications by also covering papers indexed in the WoS under "Science Citation Index Expanded".
Although we recognize the limitations of our analysis, the main aim of this paper was to analyze the academic structure of the papers, authors, institutions and countries that have cited Aitchison's 1982 paper. We believe it does this in a sufficiently rigorous and complete manner, while also presenting an overview of the most important data related to Aitchison's 1982 paper, which has recently celebrated the 35th anniversary since its publication.

Acknowledgments

The article was supported by the Spanish Ministry of Science, Innovation and Universities (grant RTI2018-095518-B-C21). Acknowledgements are due to Germà Coenders for his comments on previous drafts. We also would like to thank the anonymous reviewers for valuable comments that have improved the quality of the paper.

References

Aitchison J (1982). "The Statistical Analysis of Compositional Data." *Journal of the Royal Statistical Society: Series B (Methodological)*, **44**(2), 139–160.

Aitchison J (1986). *The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability.* Chapman & Hall (Reprinted in 2003 with additional material by The Blackburn Press).

Aitchison J (1994). "Principles of Compositional Data Analysis." *Lecture Notes-Monograph Series*, **24**, 73–81.

Bar-Ilan J (2008). "Informetrics at the Beginning of the 21st Century A Review." *Journal of Informetrics*, **2**(1), 1–52.

Batista-Foguet JM, Ferrer-Rosell B, Serlavós R, Coenders G, Boyatzis RE (2015). "An Alternative Approach to Analyze Ipsative Data. Revisiting Experiential Learning Theory." *Frontiers in Psychology*, **6**, 1742.

Belles-Sampera J, Guillen M, Santolino M (2016). "Compositional Methods Applied to Capital Allocation Problems." *Journal of Risk*, **19**, 15–30.

Blasco-Duatis M, Coenders G, Saez M, García NF, Cunha IF (2019). "Mapping the Agenda-setting Theory, Priming and the Spiral of Silence in Twitter Accounts of Political Parties." *International Journal of Web Based Communities*, **15**(1), 4–24.

Boyack KW, Klavans R (2014). "Creation of a Highly Detailed, Dynamic, Global Model and Map of Science." *Journal of the Association for Information Science and Technology*, **65**(4), 670–685.

Carreras Simó M, Coenders G (2020). "Principal Component Analysis of Financial Statements: A Compositional Approach." *Revista de Métodos Cuantitativos para la Economía y la Empresa*, **29**, 18–37.

Chayes FA (1948). "A Petrographic Criterion for the Possible Replacement Origin of Rocks." *American Journal of Science*, **246**(7), 413–425.

Coenders G, Ferrer-Rosell B (2020). "Compositional Data Analysis in Tourism: Review and Future Directions." *Tourism Analysis*, **25**(1), 153–168.

Comas-Cufí M, Thió-Henestrosa S, Egozcue J, Tolosana-Delgado R, Ortego M (2011). "CoDaWork11: 4th International Workshop on Compositional Data Analysis." *Sant Feliu de Guíxols*.

Consultancy SR (2019). "QS World University Ranking 2019."

Creixans-Tenas J, Coenders G, Arimany-Serrat N (2019). "Corporate Social Responsibility and Financial Profile of Spanish Private Hospitals." *Heliyon*, **5**(10), e02623.

Ding Y, Rousseau R, Wolfram D (2016). *Measuring Scholarly Impact*. Springer.

Egozcue JJ, Pawlowsky-Glahn V (2019). "Compositional Data: The Sample Space and Its Structure." *TEST*, **28**(3), 599–638.

Egozcue JJ, Pawlowsky-Glahn V, Gloor GB (2018). "Linear Association in Compositional Data Analysis." *Austrian Journal of Statistics*, **47**(1), 3–31.

Ferrer-Rosell B, Coenders G, Martínez-Garcia E (2015). "Determinants in Tourist Expenditure Composition The Role of Airline Types." *Tourism Economics*, **21**(1), 9–32.

Filzmoser P, Hron K, Templ M (2018). *Applied Compositional Data Analysis*. Springer.

Glänzel W, Thijs B (2012). "Using Core Documents for Detecting and Labelling New Emerging Topics." *Scientometrics*, **91**(2), 399–416.

Greenacre M (2017). *Correspondence Analysis in Practice*. CRC press.

Kessler MM (1963). "Bibliographic Coupling between Scientific Papers." *American Documentation*, **14**(1), 10–25.

Kogovšek T, Coenders G, Hlebec V (2013). "Predictors and Outcomes of Social Network Compositions: A Compositional Structural Equation Modeling Approach." *Social Networks*, **35**(1), 1–10.

Köseoglu MA, Sehitoglu Y, Ross G, Parnell JA (2016). "The Evolution of Business Ethics Research in the Realm of Tourism and Hospitality." *International Journal of Contemporary Hospitality Management*, **28**, 1598–1621.

Linares-Mustarós S, Coenders G, Vives-Mestres M (2018). "Financial Performance and Distress Profiles. From Classification according to Financial Ratios to Compositional Classification." *Advances in Accounting*, **40**, 1–10.

Lotka AJ (1926). "The Frequency Distribution of Scientific Productivity." *Journal of the Washington Academy of Sciences*, **16**(12), 317–323.

Martín-Fernández JA, Olea-Meneses RA, Pawlowsky-Glahn V (2001). "Criteria to Compare Estimation Methods of Regionalized Compositions." *Mathematical Geology*, **33**(8), 889–909.

Martyn J (1964). "Bibliographic Coupling." *Journal of Documentation*, **20**(4), 236–236.

McCain KW (1986). "Cocited Author Mapping as a Valid Representation of Intellectual Structure." *Journal of the American Society for Information Science*, **37**(3), 111–122.

McCain KW (1990). "Mapping Authors in Intellectual Space: A Technical Overview." *Journal of the American Society for Information Science*, **41**(6), 433.

McCain KW (1991). "Mapping Economics through the Journal Literature: An Experiment in Journal Cocitation Analysis." *Journal of the American Society for Information Science*, **42**(4), 290.

Merigó JM, Blanco-Mesa F, Gil-Lafuente AM, Yager RR (2017). "Thirty Years of the International Journal of Intelligent Systems: A Bibliometric Review." *International Journal of Intelligent Systems*, **32**(5), 526–554.

Merigó JM, Gil-Lafuente AM, Yager RR (2015). "An Overview of Fuzzy Research with Bibliometric Indicators." *Applied Soft Computing*, **27**, 420–433.

Morais J, Thomas-Agnan C, Simioni M (2018). "Interpretation of Explanatory Variables Impacts in Compositional Regression Models." *Austrian Journal of Statistics*, **47**(5), 1–25.

Mulet-Forteza C, Genovart-Balaguer J, Mauleon-Mendez E, Merigó JM (2019a). "A Bibliometric Research in the Tourism, Leisure and Hospitality Fields." *Journal of Business Research*, **101**, 819–827.

Mulet-Forteza C, Genovart-Balaguer J, Merigó JM, Mauleon-Mendez E (2019b). "Bibliometric Structure of IJCHM in Its 30 Years." *International Journal of Contemporary Hospitality Management*, **31**(12), 4574–4604.

Mulet-Forteza C, Salvá AS, Monserrat S, Amores A (2020). "80th Anniversary of Pure and Applied Geophysics: A Bibliometric Overview." *Pure and Applied Geophysics*, **177**(2), 531–570.

Ortells R, Egozcue JJ, Ortego MI, Garola A (2015). "Relationship between the Popularity of Key Words in the Google Browser and the Evolution of Worldwide Financial Indices." In *International Workshop on Compositional Data Analysis*, pp. 145–165. Springer.

Palarea-Albaladejo J, Martín-Fernández JA (2015). "zCompositionsR Package for Multivariate Imputation of Left-censored Data under a Compositional Approach." *Chemometrics and Intelligent Laboratory Systems*, **143**, 85–96.

Pawlowsky-Glahn V, Buccianti A (2011). *Compositional Data Analysis: Theory and Applications*. John Wiley & Sons.

Pawlowsky-Glahn V, Egozcue JJ (2001). "Geometric Approach to Statistical Analysis on the Simplex." *Stochastic Environmental Research and Risk Assessment*, **15**(5), 384–398.

Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2015). *Modeling and Analysis of Compositional Data*. John Wiley & Sons.

Pearson K (1897). "Mathematical Contributions to the Theory of Evolution on a Form of Spurious Correlation which May Arise when Indices Are Used in the Measurement of Organs." *Proceedings of the Royal Society of London*, **60**, 489–498.

Price DJ (1986). *Little Science, Big Science... and Beyond*. Columbia University Press.

Pritchard A (1969). "Statistical bibliography or bibliometrics." *Journal of Documentation*, **25**(4), 348–349.

Ramos-Rodríguez AR, Ruíz-Navarro J (2004). "Changes in the Intellectual Structure of Strategic Management Research: A Bibliometric Study of the Strategic Management Journal, 1980–2000." *Strategic Management Journal*, **25**(10), 981–1004.

Rodrigues LA, Daunís-I-Estadella J, Mateu-Figueras G, Thió-Henestrosa S (2011). *Flying in Compositional Morphospaces: Evolution of Limb Proportions in Flying Vertebrates*. Wiley Online Library.

Sinkovics N (2016). "Enhancing the Foundations for Theorising through Bibliometric Mapping." *International Marketing Review*, **33**(3), 327–350.

Small H (1973). "Co-citation in the Scientific Literature: A New Measure of the Relationship between Two Documents." *Journal of the American Society for Information Science*, **24**(4), 265–269.

Small H (1999). "Visualizing Science by Citation Mapping." *Journal of the American Society for Information Science*, **50**(9), 799–813.

Svensson G (2010). "SSCI and Its Impact Factors: A Prisoner's Dilemma?" *European Journal of Marketing*, **44**, 23–33.

Symonds Q (2019). "QS World University Ranking 2019."

Thió-Henestrosa S, Martín-Fernández JA (2005). "Dealing with Compositional Data: The Freeware CoDaPack." *Mathematical Geology*, **37**(7), 773–793.

Valls R (2018). *Why and How Should Geologists Use Compositional Data Analysis.* Valls Geoconsultant.

Van den Boogaart KG, Tolosana-Delgado R (2013). *Analyzing Compositional Data with R.* Springer.

Van Eck NJ, Waltman L (2010). "Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping." *Scientometrics*, **84**(2), 523–538.

Vistelius AB, Sarmanov OV (1961). "On the Correlation between Percentage Values: Major Component Correlation in Ferromagnesium Micas." *The Journal of Geology*, **69**(2), 145–153.

Young H (1983). *The ALA Glossary of Library and Information Science.* American Library Association.

Yuan Y, Gretzel U, Tseng YH (2015). "Revealing the Nature of Contemporary Tourism Research: Extracting Common Subject Areas through Bibliographic Coupling." *International Journal of Tourism Research*, **17**(5), 417–431.

Zupic I, Čater T (2015). "Bibliometric Methods in Management and Organization." *Organizational Research Methods*, **18**(3), 429–472.

Affiliation:

Carolina Navarro-Lopez
Department of Economics
Faculty of Tourism
University of the Balearic Islands
E-mail: carolina.navarro@uib.es
URL: https://www.uib.eu/personal/ABTE2NjI2/

Silvia Gonzalez-Morcillo
Department of Economics
Faculty of Tourism
University of the Balearic Islands
E-mail: s.gonzalez-morcillo1@estudiant.uib.eu

Carles Mulet-Forteza
Department of Economics
Faculty of Tourism
University of the Balearic Islands
E-mail: carles.mulet@uib.es
URL: https://www.uib.eu/personal/ABTE1NDI1/

Salvador Linares-Mustaros
Department of Business Administration
University of Girona
E-mail: salvador.linares@udg.edu
URL: https://www.udg.edu/es/directori/pagina-personal?ID=2002049

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 50 *Submitted:* 2019-11-28
January 2021 *Accepted:* 2020-07-18

8. Article #2

Original Research

SAGE Open
April-June 2022: 1–17
© The Author(s) 2022
DOI: 10.1177/21582440221093366
journals.sagepub.com/home/sgo
SAGE

"The Statistical Analysis of Compositional Data" by John Aitchison (1986): A Bibliometric Overview

Carolina Navarro-Lopez[1], Salvador Linares-Mustaros[2], and Carles Mulet-Forteza[1]

Abstract

This paper presents a complete bibliometric analysis of Aitchison's 1986 seminal book "The Statistical Analysis of Compositional Data." We have set three objectives. The first is to analyze the academic structure of Aitchison's 1986 book. Results reveals that although the work has received citations uninterruptedly since its publication, the number of these has increased very significantly over the past 4 years. This is due to the significant increase in the number of publications on the theme of "Compositional Data Analysis" in fields related to "geoscience" over the last few years. The second objective is to determine which main journals Aitchison's book has been cited in. The results highlight that the main journals are indexed under the following WoS category: "Geosciences, Multidisciplinary" and "Ecology." Of these, "Mathematical Geosciences" and "Computers, Geosciences" stand out. The third objective is to determine the main topics analyzed in the principal papers published by authors citing Aitchison's book. Our results show that the keywords in the main papers to have cited Aitchison's 1986 book originate from the geoscience field, since many of them are related to concepts directly linked to this field and refer to terms related to "biodiversity," "geodiversity," "geoheritage," and "georesources." Lastly, the analysis shows how the CoDA methodology is now in a phase of exponential growth, expanding to other fields. This implies that geoscience is becoming consolidated in the scientific literature as one of the branches of modern science that has given rise to a new mathematical theory of great impact.

Keywords

John Aitchison, bibliometrics, citation structure, journal ranking, hot topics, VOSviewer

Introduction

Compositional data are positive data that carry only relative information and in the most common situations they sum up to a constant (Filzmoser & Hron, 2009). They are frequent in geology and chemistry, for example, since total amounts are trivially related to the size of the soil or chemical sample, so that only relative importance is of interest. Using standard statistical techniques with compositional data can produce inconsistent results due to a set of undesirable problems, including the problem of spurious correlation of ratios (Pearson, 1897), dependency on scale, appearance of outliers and asymmetry, out-of-range forecasts (negative or above the constant sum) or inconsistency of the sub-composition (Aitchison, 1986). The lack of a solution to the problems inherent in compositional data led Miesch (1969) to state that the problem of the constant sum was one of the most important and most difficult problems encountered when analyzing and interpreting geochemical data. In 1986, Aitchison presented a book entitled The Statistical Analysis of Compositional Data (Aitchison, 1986), which detailed a whole set of techniques based on compositional data, the results obtained being consistent when based on solid mathematical foundations (e.g., Daunis-i-Estadella et al., 2011; Graffelman et al., 2018; Thomas & Aitchison, 2005; Tolosana-Delgado et al., 2019; Verma et al., 2006). At present, despite the proliferation of other manuals analyzing the foundations postulated in Aitchison's work (see, e.g., Filzmoser et al., 2018; Greenacre, 2018; Pawlowsky-Glahn et al., 2015; Van Den Boogaart & Tolosana-Delgado, 2013), said foundations are still valid, and provide a solid basis for

[1]University of the Balearic Islands, Ctra. Valldemossa, Palma de Mallorca, Spain
[2]University of Girona, Catalunya, Spain

Corresponding Author:
Carles Mulet-Forteza, Department of Business Administration, University of the Balearic Islands, Edificio Jovellanos, Ctra. Valldemossa, Km 7, 5, Mallorca 07122, Spain.
Email: carles.mulet@uib.es

validating results, which is why Aitchison's (1986) book is universally considered to be both essential and seminal.

Consolidation of the statistical methods compiled in Aitchison's book for the field of geoscience (Chakraborty et al., 2020; Hron et al., 2021; Mikšová et al., 2020; Pawlowsky-Glahn & Egozcue, 2020; Pospiech et al., 2021) and the current expansion of these techniques to new scientific fields, such as chemistry, biology, medicine, psychology, education, communication, demography, geography, and other social science disciplines (Batista-Foguet et al., 2015; Belles-Sampera et al., 2016; Blasco-Duatis et al., 2018; Carreras Simó & Coenders, 2020; Carreras-Simó & Coenders, 2021; Coenders & Ferrer-Rosell, 2020; Ezbakhe & Pérez Foguet, 2020; Ferrer-Rosell et al., 2015; Kogovšek et al., 2013; Linares-Mustarós et al., 2018; Muller et al., 2018; Ortells et al., 2016; Rodrigues et al., 2011; Sanz-Sanz et al., 2018) explains the exponential growth in the number of citations of the book, while at the same time confirming the claim that geoscience is establishing itself in the scientific literature as a branch of modern science that has adopted a new mathematical theory of great impact.

This paper presents a bibliometric analysis of the aforementioned book, "The Statistical Analysis of Compositional Data," with the aim of studying the relationships existing between this seminal publication in the field of geoscience and modern science. In the first part of this analysis, we present an overview of the academic structure of publications that have cited Aitchison's work, while in the second part we present a study of the main journals and research topics to have been addressed in articles by authors who have cited it, with the aim of providing an answer to the following research questions (RQ):

RQ1: What is the academic structure of Aitchison's 1986 book?
RQ2: In which main journals has Aitchison's book been cited?
RQ3: What main topics are analyzed in the principal papers published by authors citing Aitchison's book?

This paper makes several important contributions. First, the bibliometric analysis contributes to the growing literature of articles summarizing the achievements and trends in research fields over long periods of time. Identifying the citation structure, origins and evolution of the main topics addressed, as well as the main sources used by authors citing Aitchison's research, will help us to determine intellectual connections in academic fields that use CoDA in their field of research (Koseoglu et al., 2019; Köseoglu et al., 2019; Shafique, 2013). In respect of this, mapping intellectual connections aids the creation of new theories and the development of existing theories, providing a glimpse of future directions that scientific research may take (Köseoglu et al., 2021). Thus, conducting an academic analysis of such developments can help researchers identify the potential impacts theories may have on society. In addition, these processes also provide valuable information for both academics and practitioners (Torraco, 2016), by providing them with a study that helps to have an outline of the status quo of CoDA, especially for those who are not very familiar with CoDA but are interested in it (Jiang & Fan, 2022). Second, the bibliometric study provides a comprehensive picture of specific research fields and allows researchers to focus on unique areas to add new results and knowledge to the literature (Ghorbani et al., 2021). Then, this paper contributes to a better understanding of the current status, development and future lines of research in the field, supporting researchers and other experts in identifying research areas, and selecting the most appropriate journals to publish their own findings (Sajovic & Boh Podgornik, 2022). Thirdly, the research conducted delves into intellectual connections across a large body of research covering various fields related to geoscience. Fourthly, the work encompasses a long time horizon, allowing researchers to obtain a complete picture of the field addressed, as well as its evolution. Fifthly, the document citation networks are analyzed and reference citation bursts detected in order to provide information on research topics and assess trends over time from different perspectives, which will be of great use for future research; in other words, this study helps to provide an orchestration of knowledge in the field. And finally, the present work focuses on documents that have passed the strict refereeing process, meaning that the results obtained are highly reliable.

Bibliometrics and Social Network Analysis

Bibliometrics entails the quantification of academic production based on certain classifications that project indirect indications on its perception (Huang et al., 2019). Multiple definitions of the term exist, although the modern version is usually attributed to Alan Pritchard. Pritchard (1969) defined bibliometrics as "the application of mathematical and statistical methods to books and other means of communication." More recently, other authors have provided further definitions, however. For example, Zupic and Čater (2015) posited that bibliometrics constitutes a tool for evaluating the evolution of research areas based on social, intellectual, and conceptual structures. Therefore, we can assume bibliometrics is a discipline that aims to evaluate and map scientific progress through classification using statistical techniques (Diodato, 1994; Jappe, 2020; McBurney & Novak, 2002).

As for methodology, bibliometrics focuses exclusively on measuring publications. However, the term "publication" is relatively ambiguous, since, among other documents, it may include book chapters, journal articles, and proceedings in conference volumes. Therefore, before starting a bibliometric research project, it is important to clearly define what is being measured and what type of publication should serve as the basis for the bibliometric analyses to be carried out, since bibliometrics should provide information about all the key components of a research project.

Bibliometric analysis is a fundamental statistical instrument for analyzing the state of knowledge in a given scientific area, given that it measures the number of documents published and the number of citations received for those documents. In addition, bibliometrics allows the results of the analysis to be mapped through spatial visualization of the findings with respect to the structure and dynamics of scientific fields (Boyack & Klavans, 2014; Zyoud & Fuchs-Hanusch, 2020). Its main objective is to create a representation of the network structure of a research field that highlights the connections between the main journals, publications, etc. and the topics and other key features of the analyzed field (Bruns et al., 2020; Gumpenberger et al., 2012; Vogel, 2014).

A further aim of bibliometrics is to evaluate the quality of research (Bornmann & Leydesdorff, 2014; Segura-Robles et al., 2020). At present, two main methods are used to this end: a qualitative (by peers) and a quantitative review (bibliometrics) (Feng, 2020). In this respect, the former includes particular, non-quantifiable evaluations made by experienced experts, while the latter considers a publication to be more relevant the more citations it receives.

Today, new alternatives to the classic "citation" have emerged to assess the importance of a scientific document, such as libmetrics and altimetrics. Libmetrics establishes a connection between the importance of a scientific article or book and its availability in a library by measuring how often it is acquired or borrowed from the library, for example. Altimetrics generates new knowledge by combining all of the data available online and applying big data technologies. This allows for bibliometric approaches to focus on correlations rather than causalities, since it should permit the analysis of new connections that have not previously been weighted or questioned. These alternative bibliometric methods are based on free online content, most of which is taken from social networks, which complement the data offered by bibliometrics based on conventional databases, such as the Web of Science (WoS) or Scopus.

Literature Review

Analyzing an academic discipline or scientific field is common practice nowadays, since it helps researchers develop new theories and journal editors foresee research trends (Gatrell & Breslin, 2017; Post et al., 2020; Torraco, 2016; Webster & Watson, 2002). There is therefore a growing interest in and demand for investigations into the intellectual structure of research areas or scientific fields in order to highlight progress in this regard (Kunisch et al., 2018; Torma & Thøgersen, 2021).

Bibliometrics, the main objective of which is to measure scientific output (Wang et al., 2019), emerged in the early 20th century, when psychologists began to collect statistics on publications related to their field of research (Godin, 2006). However, it was the exponential growth in academic publications in the 1950s that first saw American chemist Eugene Garfield begin to evaluate and carry out systematic counts of publications based on the literature used and cited.

One application of bibliometric methods is their use as a tool to evaluate any research that has been conducted (Bornmann & Leydesdorff, 2014; Karakus et al., 2021; Moral-Muñoz et al., 2020). This is the easy part of bibliometrics, since it provides direct information and does not require assumptions for its production. Trying to assess the quality and importance of published papers is a much more complex and less obvious task, however. Researchers have essentially used two methods to carry out this type of analysis of an academic field. The first comprises a qualitative evaluation by researchers (Lopes & Martins, 2021; Zupic & Čater, 2015). This method has several drawbacks, among which can be highlighted its subjectivity and a lack of transparency, which negatively impacts on its reliability and validity (Cook et al., 1997; Szomszor et al., 2021). The second is to conduct bibliometric analyses, and more specifically, analyzing the co-citation of documents (Lopes & Martins, 2021; Zupic & Čater, 2015), which entails identifying the intellectual structure of a scientific field by means of mathematical and statistical methods (Culnan, 1986; Hou et al., 2018). This second approach is the one most used by researchers (Hota et al., 2020; Lampe et al., 2019; Zhao et al., 2018), since it allows the tracking of practically all aspects of scientific collaboration networks (Vasilyeva et al., 2021; Ye et al., 2013).

We can therefore state that bibliometrics is a discipline that aims to assess and map the progress made in scientific fields through the classification of data. This entails, among other methods, the use of statistical techniques to analyze research performance by individuals, institutions, countries, mapping the structure of the analyzed field, etc. (Karakus et al., 2021).

The discipline has since evolved and is now used to evaluate the impact of publications, journals, authors and institutions in order to determine patterns of influence (Biemans et al., 2010; Clark et al., 2014; Post et al., 2020; Sarin et al., 2018).

Bibliometric documents have expanded into several fields (Butt et al., 2021), including accounting (Merigó & Yang, 2017), computer science (Chen et al., 2020; Garousi & Fernandes, 2017), energy (Liu et al., 2020), ecology (Jankó et al., 2017; Zhang et al., 2017), health care sciences services (He, Fang, Chen, et al., 2020; He, Fang, Wang, et al., 2020), hospitality (García-Lillo et al., 2016), medicine (Fan et al., 2020), tourism (Mulet-Forteza et al., 2019), and social media (Leung et al., 2017).

There are also bibliometric works that, rather than focusing on a specific field, analyze the publications of a particular country, institution or author. Thus, for example, Salisu and Salami (2020) analyzed publications by Nigerian authors between 1901 and 2016, Ahmad et al. (2020) analyzed the performance of publications by the University of the Punjab, and Haustein and Peters (2020) analyzed the publications made by the researcher Judit Bar-llan. These are just a few examples of these types of bibliometric analyses. To the best of our knowledge, however, no bibliometric work has yet been carried out that focuses on one specific source in the bibliometric literature, making this paper a starting point for future bibliometric studies.

The results of bibliometric studies provide very useful information for policymakers and academic decision-makers in universities, research centers and governments, as they are considered reliable and relevant sources of results, and are often used to justify decisions on research policies, job offers and promotions, as well as to direct and support research projects (Bornmann & Leydesdorff, 2014; Gatrell & Breslin, 2017; Gläser & Laudel, 2015; Post et al., 2020). In addition, both public and private research funding agencies often ask researchers to either provide certain indications of quality to fund their research or to demonstrate that the research to be carried out has the potential to impact society (Bornmann, 2014; Brueton et al., 2014; Smits & Champagne, 2020). By way of example, in the United Kingdom bibliometrics has been considered for assessing the quality of research output within the country's framework for research excellence. Finally, we would like to point out that bibliometrics can also help journal editors evaluate past publications, design new policies and make future editorial decisions.

Methodology

The statistical data used in this paper were compiled from the WoS database in November 2019. According to Merigó et al. (2015), the WoS comprises information from over 15,000 sources and 50,000,000 documents ranked according to over 250 categories and 150 research areas. It is widely considered to be the most influential in the world.

The bibliometric data used in this work were obtained as follows. First, the "Cited Reference Search" option in the WoS database was used. Subsequently, the "Cited work" option was selected and the following text entered: "The Statistical Analysis of Compositional Data." The search yielded a total of 69 records, which were reduced to 58 once the records that do not make specific reference to the paper "The Statistical Analysis of Compositional Data by Aitchison (1986)" were eliminated. The 11 deleted records, seven primarily referred to another work by Aitchison (1982), in which he first introduced the concept of Compositional Data Analysis, although in a much shorter form and with less impact than the work published in 1986. The remaining four deleted records referred to another work by John Aitchison entitled "Logratios and natural laws in compositional data analysis," published in the journal Mathematical Geology in 1999. Once these 11 records had been eliminated, the remaining 58 were selected, which do make explicit reference to the paper "The Statistical Analysis of Compositional Data by Aitchison (1986)." These 58 records returned a total of 2,636 papers that had cited Aitchison's 1986 book. Finally, the number of documents was reduced to 2,426 after limiting the search to only those that had passed a strict arbitration process, including papers, reviews and letters (Merigó et al., 2019).

In the following step, the option "Cited work" was selected and the following text entered: "The Statistical Analysis of Compositional Data." In this paper, we have considered a wide range of bibliometric methods to represent the bibliographical data analyzed. First, we considered the number of publications and citations, which are the most popular methods according to Ding et al. (2014). Whereas the number of citations generally measures influence, productivity is measured by the number of documents (Svensson, 2010). Another indicator we used here refers to the most influential keywords (Mulet-Forteza et al., 2019).

The VOSviewer Software was used to map the consulted bibliographic data for co-occurrence of author keywords and co-citations (Van Eck & Waltman, 2010, 2014). Such maps allow several aspects of a scientific field to be monitored (Noyons et al., 1999; Su et al., 2019), providing a clearer view of the results obtained (Merigó et al., 2016). Keyword co-occurrence refers to the most common keywords used to develop a research field or a scientific document (Callon et al., 1983; Ding et al., 2001; Huang et al., 2019; McCain, 1986, 1991; Zhang et al., 2019), while co-citation assumes that there is some kind of association between two documents jointly cited by a different third one (Boyack & Klavans, 2014; Hoque et al., 2021; McCain, 1990; Ramos-Rodríguez & Ruíz-Navarro, 2004; Small, 1973).

Before we could perform the graphical analysis with the VOSviewer Software, we had to clean up the data collected from the WoS. In order to carry out the co-citation analysis of journals, the names of the journals with different designations had to be unified. By way of example, data appearing under the names "j roy stat soc b" and "j roy stat soc b m," those appearing under the names "soil sci" and "soil sci s," or those appearing under the names "behav ecol" and "behav ecol s" were unified under the same name. Journals that changed their names during the period were also unified, such as "*Mathematical Geology*," which changed its name to "*Mathematical Geosciences*" *in 2008*.

The same had to be done in the co-occurrence of author keyword analysis. In this case, keywords that appeared simultaneously in the singular and plural, such as "stream sediment" and "stream sediments" or "ternary diagram" and "ternary diagrams," keywords that appear with or without a hyphen, such as "particle size distribution" and "particle-size distribution" or "isometric log-ratio transformation" and "isometric logratio transformation," and keywords that are written differently in American and British English, such as "foraging behavior" and "foraging behavior," had to be cleaned.

The combination of methods used in this paper allowed us to collect data using "full counting" and "fractional counting" methods. With the former, a publication co-authored by several researchers is assigned to each researcher with a full weight of one, while the "fractional counting" method (VOSviewer software) divides the authorship of the

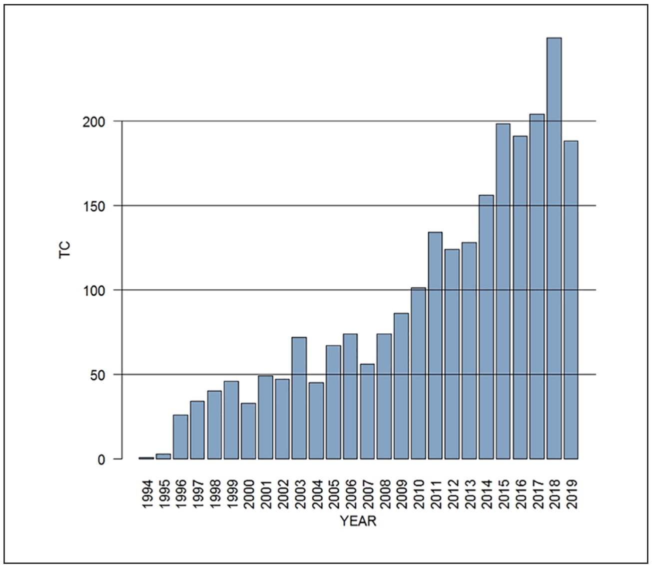

Figure 1. Annual number of citations received by Aitchison's 1986 book.
Source. Authors, WoS database, 1986 through November 2019.
Note. TC = total citations.

document among the number of authors (Mulet-Forteza et al., 2019). In this regard, it should be borne in mind that developing bibliometric networks is not a trivial process and, depending on how this is done, they can yield very different results, as Perianes-Rodriguez et al. (2016) showed for the case of journal network analysis. These authors argued in favor of the "fractional counting" method for producing bibliometric maps of journals, based on the fact it awards the same influence to each reference cited in a publication.

Thus, they considered it more reasonable to use analyses based on the idea of treating each reference cited in a publication as equally representative, as is the case using the aforementioned "fractional counting" method. Although this justification seems plausible to us, as far as we are aware, the reality is that researchers have traditionally preferred to use the "full counting" method in their bibliometric map analysis. That being said, it is not our intention to take a position in favor of one method or the other here, and we therefore provide the results using both methods, which will allow us to compare the results obtained by both systems of data collection.

Bibliometric Study of Aitchison's (1986) Book

We will now address the different research questions posed in our paper.

Academic Structure of Aitchison's (1986) Book

Regarding the first question (RQ1), Figure 1 presents the evolution of citations received by Aitchison's 1986 book.

Table 1. Sections of a documents in which the Aitchison's (1986) book has been cited.

Sections of a document	TP
Introduction	229
Literature review	535
Methodology	688
Results and discussion	305
Conclusion	183
Total	1,940

Source. Authors.
Note. The same document can cite the Aitchison's (1986) book in two or more sections. TP = total papers.

The above figure shows how Aitchison's 1986 book entitled "The statistical analysis of compositional data" has received citations uninterruptedly since 1994. It also shows how different periods can be distinguished for the number of citations received. Thus, between 1994 and 2007, with ups and downs, the average annual number of citations remained at around 42. However, from 2008 onward, the annual number of citations generally increased each year, with notable jumps in 2011 and 2015. This increase in the number of citations received by the book is due, among other aspects, to the significant increase in the number of publications on the theme of "Compositional Data Analysis" in fields related to "geoscience" over the past few years and is likely to continue in the future.

We have delved into the reasons for citing Aitchison's 1986 book. First of all, we downloaded the documents citing this reference. In this regard, we would like to point out that we only had access to 1,529 of the 2,426 references citing Aitchison's 1986 book, which represents 63% of the total. The remaining 897 references could not be analyzed because the databases of the universities of the different authors who have written this document do not have access to all the journals. Even so, we consider that we have analyzed a significant percentage of references that validate the comments made above.

Table 1 shows where Aitchison's 1986 book has been cited in the paper.

Table 1 shows how most of the documents citing Aitchison's 1986 book have used CoDA in their methodologies. Specifically, 35% of the documents analyzed cite Aitchison's 1986 book in the methodology, while 16% do so in the results and discussion section. It is also noteworthy that 28% of the analyzed papers cite it in the literature review, while only 12% and 9% cite it in the introduction and conclusion, respectively.

It is also interesting to carry out a temporal analysis to determine if there is a period of time after which CoDA methodology has started to be used effectively in the papers (Table 2).

Table 2 shows how during the first period analyzed (1994–2000) the citations obtained by Aitchison's 1986 book were

Table 2. Sections of a document in which the Aitchison's (1986) book has been cited. Temporal evolution.

	1994–2000	2001–2010	2011–2019
Introduction	35	55	139
Literature review	28	169	338
Methodology	24	245	419
Results and discussion	18	102	185
Conclusion	16	71	96

Source. Authors.
Note. The same document can cite the Aitchison's (1986) book in two or more sections.

concentrated in the introduction of the documents that cite it. On the other hand, during the period 2001 to 2010 it can be seen how these are distributed, in percentage terms, in a similar way between the literature review, methodology and results, and discussion sections, although the conclusion section is the one which, in percentage terms, has the highest number of citations. Finally, during the period 2011 to 2019 it can be observed, also in percentage terms, how the conclusion section loses weight when citing Aitchison's 1986 book, while the rest of the sections increase their percentage when citing this document. All this shows that, during the last period analyzed, the CoDA is analyzed both from a literature review and from the use of this methodology, which indicates that this technique already enjoys a notable maturity and scientific applicability.

Main Journals Citing Aitchison's (1986) Book

In this section, we will address the second question (RQ2) posed in our paper. Firstly, Table 3 shows the main journals to have most cited Aitchison's (1986) book.

Table 3 shows how most of the documents that have cited Aitchison's (1986) book are published in the *Journal of Geochemical Exploration* and *Mathematical Geosciences*, followed by *Plos One* and *Applied Geochemistry*. Table 3 also shows how the aforementioned journals, together with *Ecology*, *Geochimica et Cosmochimica Acta*, *Journal of Chemical Ecology*, *Chemical Geology*, and *Evolution International Journal of Organic Evolution* are the ones that present the most important strength of connections. In this regard, Figure 2 provides further details of the 500 most important connections occurring between the journals that have cited the book. Table 3 also shows that nine journals on the list are directly related to the field of "geoscience." Specifically, here we are referring to the journals "Mathematical Geosciences," "Computers Geosciences," "Archaeometry," "Journal of Archaeological Science," "Palaeogeography Palaeoclimatology Palaeoecology," "Catena," "Quaternary International," "Quaternary Science Reviews," and "Journal of Quaternary Science."

A map was conducted to reflect the main relationships established between the journals citing the book. In addition, Figure 2 shows the 500 main co-citation links between the principal journals citing the book.

Figure 2 reveals six main clusters, each represented by the same color. Larger clusters include a greater number of journals that have cited Aitchison's book. The distance between two clusters shows the relationship of the clusters in terms of citations, where the clusters located close to each other tend to be related, and vice versa. Within one cluster, the size of a circle represents the number of times a journal has cited the book, larger circles therefore indicating journals that have cited it a greater number of times. The thickness of the curved lines between the clusters represents the number of citations between two journals, whether they belong to the same cluster or not. And finally, the name of each circle (or label) indicates the name of the journal. In this regard, it should be noted that the VOSviewer Software aims to avoid overlapping labels, meaning that the labels are not visible for some journals in Figure 2. The above description also applies to Figures 3 to 6.

The first cluster in Figure 2, in red, comprises 95 journals indexed mainly under the WoS categories "Geochemistry & Geophysics" and "Chemistry." What the research carried out in these categories has in common, among other aspects, is that it usually considers a large number of variables in its analysis. In these cases, the CoDA methodology improves the results obtained from these analyses, as the proportionality features of abundance data are fully taken into account, thereby enhancing their relative multivariate behavior (Buccianti et al., 2015). These particular features made these fields pioneers in applying statistical methods based on CoDA applications, especially by members of the International Association for Mathematical Geosciences. The second cluster, in green, is composed of 94 journals indexed mainly under the WoS categories "Multidisciplinary Sciences" and "Ecology." These fields are similar to those of the first cluster, as they are ones in which studies of different species abound and in which percentages are widely used to infer the ecological preferences found among species. In this case, CoDA allows for the elimination of inconsistencies that occur when determining percentages, which later become false correlations (Guerreiro et al., 2015).

The third cluster, in dark blue, comprises 54 journals indexed mainly under the "Statistics & Probability" and "Mathematics" categories. Logically, these categories form the central axis of Figure 2, CoDA applications being very useful in these fields for eliminating all kinds of mathematical and statistical inconsistencies that can be caused by working with percentages. The fourth cluster, in yellow, is composed of 45 journals indexed basically under the categories "Soil Sciences" and "Environmental Sciences." The fifth cluster, in purple, is composed of 39 journals indexed mainly under the category "Ecology." And finally, the last cluster, in blue, is composed of 33 journals indexed mainly under the category "Zoology." These last three clusters maintain certain characteristics similar to the first two clusters described above, hence their widespread use of CoDA methodology. Therefore, we observe that Aitchison's 1986 book has

Table 3. Main journals that have cited Aitchison's (1986) book.

R	Journal	TP	TLS
1	Journal of Geochemical Exploration	82	1,203.31
2	Mathematical Geosciences	68	2,074.38
3	Plos One	47	745.01
4	Applied Geochemistry	45	659.41
5	Sedimentary Geology	34	617.42
6	Journal of Chemical Ecology	33	841.81
7	Journal of Archaeological Science	27	324.76
7	Science of the Total Environment	27	748.14
9	Computers Geosciences	24	622.50
9	Evolution International Journal of Organic Evolution	24	804.62
11	Journal of Wildlife Management	23	476.07
12	Archaeometry	22	359.94
12	Journal of Applied Statistics	22	146.48
12	Journal of Evolutionary Biology	22	292.49
15	Animal Behaviour	20	657.30
15	Geoderma	20	424.32
17	Geochemistry Exploration Environment Analysis	19	386.76
18	Ore Geology Reviews	18	433.21
19	Palaeogeography Palaeoclimatology Palaeoecology	16	383.33
20	Behavioral Ecology	15	690.50
21	Catena	14	175.55
21	Geochimica et Cosmochimica Acta	14	989.86
21	Wildlife Biology	14	58.41
24	Austrian Journal of Statistics	13	77.45
24	Bird Study	13	117.12
24	Environmental and Ecological Statistics	13	62.65
24	Journal of the American Statistical Association	13	688.75
24	Proceedings of the Royal Society B Biological Sciences	13	654.84
24	Quaternary International	13	291.26
24	Stochastic Environmental Research and Risk Assessment	13	203.98
31	American Naturalist	12	686.88
32	Behavioral Ecology and Sociobiology	11	82.38
32	Environmetrics	11	192.29
32	Insectes Sociaux	11	274.49
32	Journal of Sedimentary Research	11	243.79
32	Quaternary Science Reviews	11	616.78
32	Statistical Modelling	11	84.27
38	Chemical Geology	10	824.22
38	Ecology	10	1,051.88
38	Journal of Insect Physiology	10	265.43
38	Journal of Quaternary Science	10	267.56
38	Oecologia	10	411.18

Source. Authors, WoS database, 1986 through November 2019.
Note. The records of the journals that have changed their name during the analyzed period have been unified under the most recent name of the journal, such as "Mathematical Geology" which changed, in 2008, its name to "Mathematical Geosciences." R = ranking; TP = total papers; TLS = total link strength.

received citations in a multitude of WoS categories, although most are fields related to "geoscience."

Since Figure 2 is very difficult to read, given the large number of journals appearing in it, we have produced a new one, Figure 3, which presents the same results, but mapped with a higher threshold in order to observe the journals that cite the book in greater detail.

In this case, Figure 3 is composed of five main clusters, which are further detailed in the following two figures.

Figure 3 distributes the journals among five clusters, the first consisting of 25 journals, the second 18, the third 11, the fourth ten and the third nine. This more detailed view provided by the previous three figures, especially Figures 4 and 5, allows us to take a closer look at the main

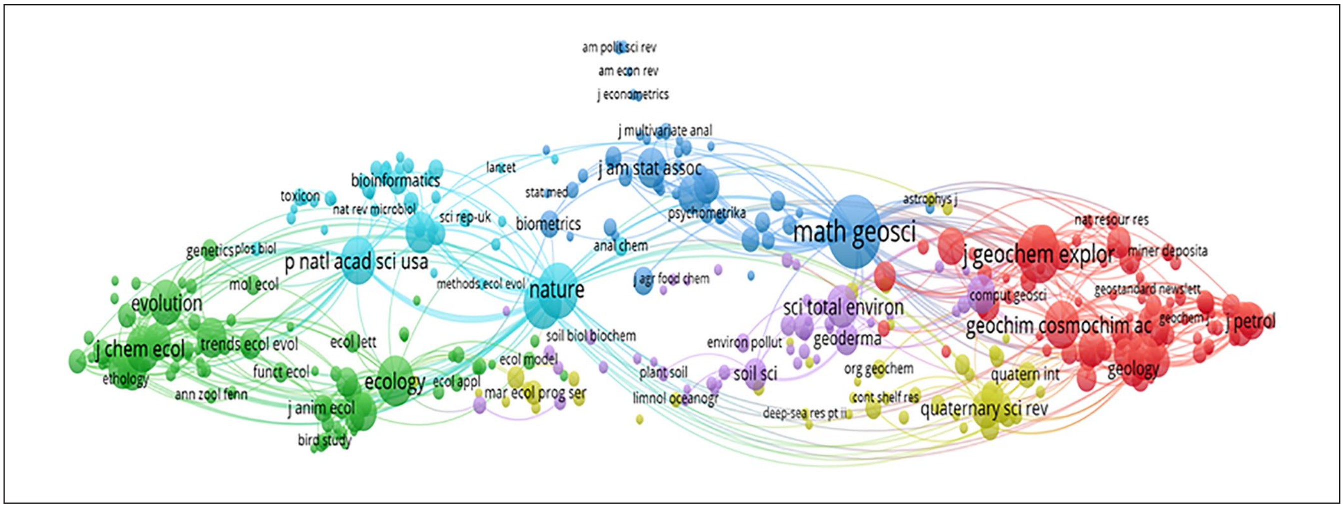

Figure 2. Co-citations of journals that have cited Aitchinson's (1986) book. Citation threshold of 50 and showing the 500 most representative co-citation connections.
Source. Authors, based on WoS database; figure created using VOSviewer Software.

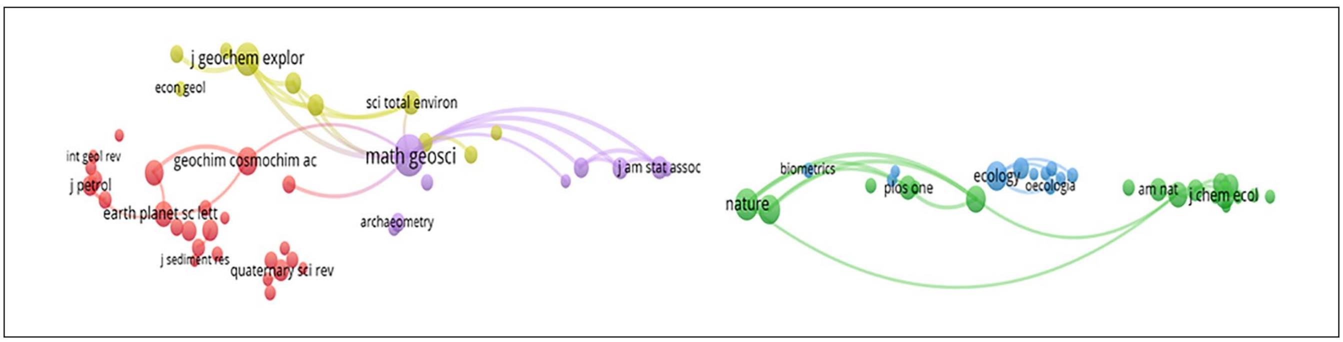

Figure 3. Co-citations of journals that have cited Aitchinson's (1986) document. Citation threshold of 250 and showing the 50 most representative co-citation connections.
Source. Authors, based on WoS database; figure created using VOSviewer Software.

journals to have published most work based on the CoDA methodology. This is naturally of great help to researchers who use this methodology in their publications, as they are able to relatively easily identify potential journals in which to publish their research, as well as ones they should consult to find out the recent directions taken by research based on the CoDA methodology in their fields of study. In this case, the categories most represented in the previous figure are as follows (in this order): "Geosciences, Multidisciplinary," "Ecology," "Geochemistry & Geophysics" and "Statistics & Probability." Specifically, 23 journals are indexed under the first category, 19 in the second, 16 in the third and nine in the fourth. This reveals how the field of "geoscience" has become the main one to use the CoDA methodology, leading to the spread of a mathematical theory of great academic impact. All of this is evident from Table 4, which was compiled using the information available in Figure 2.

Main Topics Citing Aitchison's (1986) Book

The third research question will be addressed in this section, since here we will analyze the main topics of the most relevant papers published by authors who cite Aitchison's (1986) book. Figure 6 shows a co-occurrence of keywords in the papers citing the book.

Figure 6 reads identically to the previous ones, with the following differences: in this case, the size of a term reflects the number of times the term has been cited in publications citing the book, and the distance indicates the strength of the relationship between the terms. Colors indicate groups of nearby terms in relation to co-occurrences. Finally, the strongest relationships are indicated with curved lines.

Figure 6 shows an analysis of keywords and their possible connection. Having analyzed the number of keyword occurrences, we observe that Figure 6 has nine keyword clusters. Clusters 1 and 2 are the most numerous and have

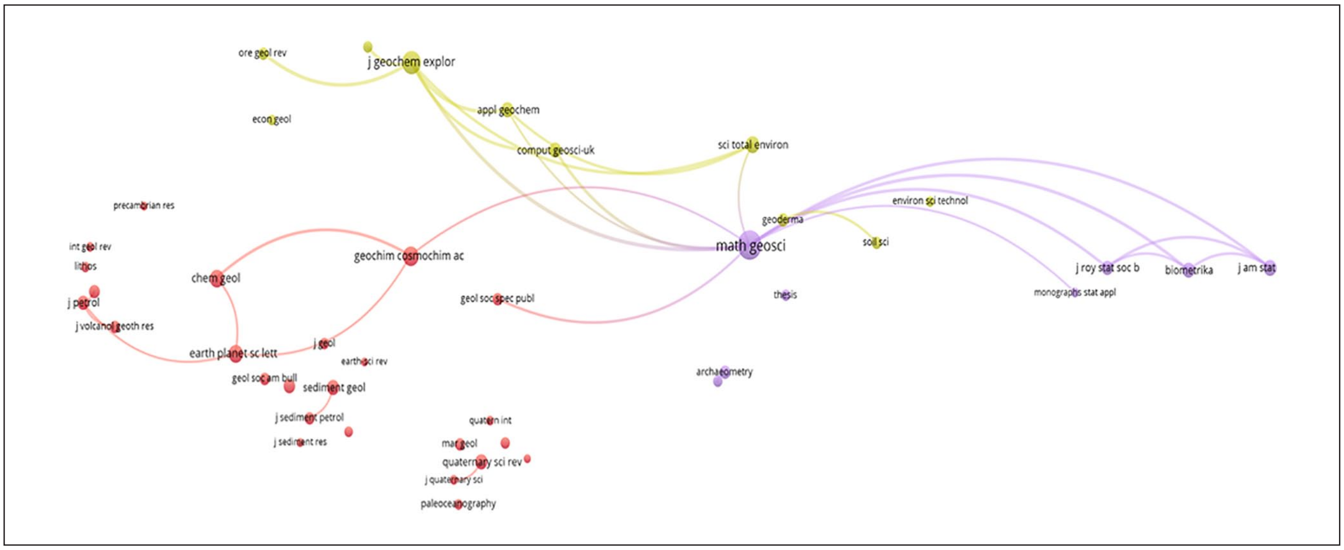

Figure 4. Co-citations of journals that have cited Aitchinson's (1986) book. Red, yellow and lilac clusters. Citation threshold of 250 and showing the 50 most representative co-citation connections.
Source. Authors, based on WoS database; figure created using VOSviewer Software.

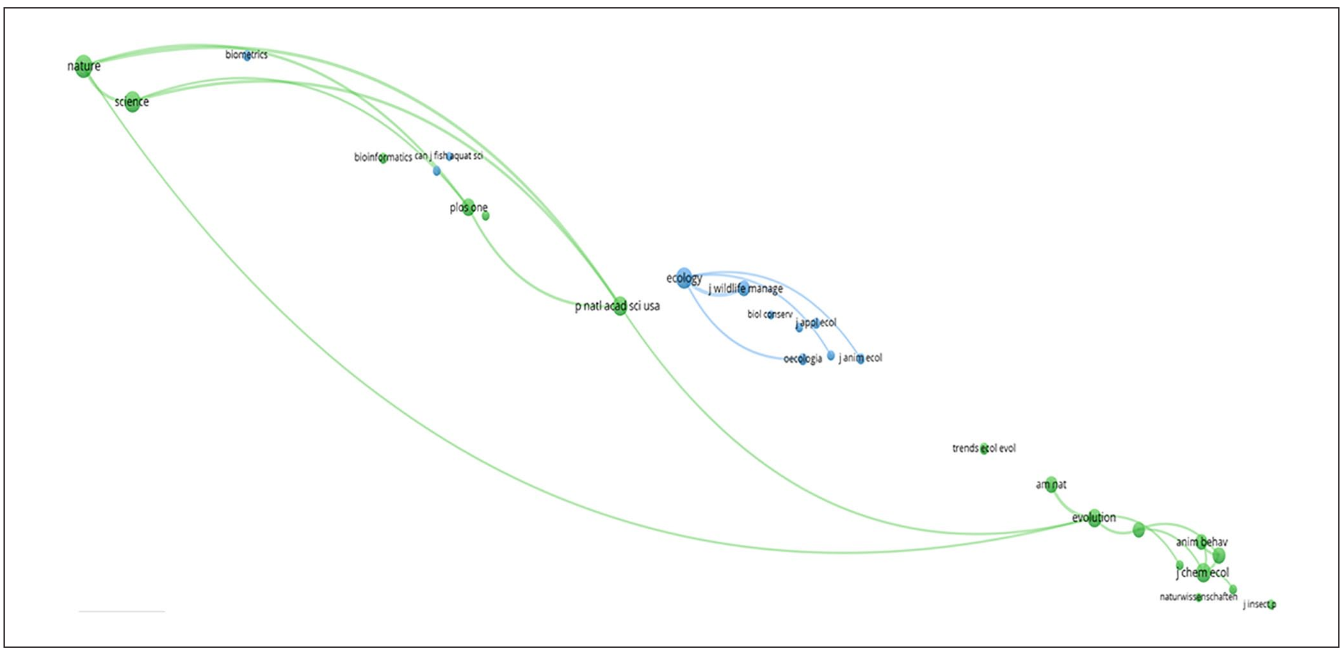

Figure 5. Co-citations of journals that have cited Aitchinson's (1986) book. Green and blue clusters. Citation threshold of 250 and showing the 50 most representative co-citation connections.
Source. Authors, based on WoS database; figure created using VOSviewer Software.

12 keywords each. In the first cluster, the words "principal component analysis," "log-ratio" and "simplex" stand out. In Cluster 2, the most important words are "soil," "microbiome" and "fatty acids." Clusters 3 and 4 also coincide in terms of number, having 10 keywords each. Cluster 3 ranks third in number of citations, and has "cuticular hydrocarbon," "sexual selection" and "mate choice" among its most important keywords. In Cluster 4, which is secondary, the words "multivariate," "cluster analysis" and "geostatistics" stand out. Although this cluster has a high number of words, the total occurrence is relatively low. As for Cluster 5, it has nine keywords, with the main words being "geochemistry," "provenance" and "statistics." Cluster 6 has five keywords; despite the smaller number of words, it becomes a core

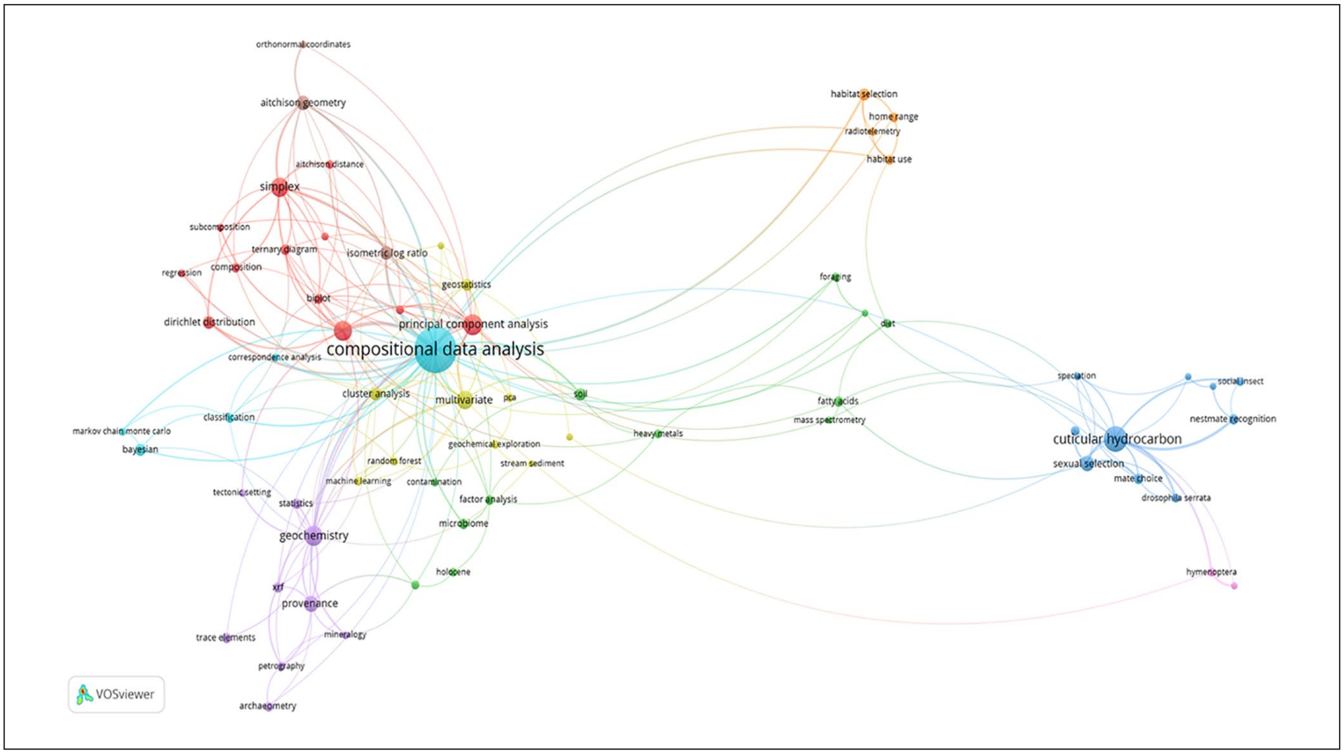

Figure 6. Co-ocurrence of keywords used by authors who have cited Aitchison's (1986) book. Citation threshold of 10 and showing the 200 most representative co-citation connections.
Source. Authors, based on WoS database; figure created using VOSviewer Software.

cluster, since it has the term with the highest number of occurrences, "compositional data analysis." As for Cluster 7, we find four keywords, with "habitat selection" the most important of these. Cluster 8 has three keywords, "Aitchinson geometry" being the most important, and finally, Cluster 9 has two keywords, the most important being "hymenoptera," although only a small difference is observed between the latter two. Figure 6 also shows how many of the keywords originate from the "geoscience" field, since many of them are related to concepts directly linked to this field and refer, in turn, to terms related to "biodiversity," "geodiversity," "geoheritage" and "georesources" (Thomas, 2016).

Since it is difficult to observe the strength of the links between the main keywords in Figure 6, Table 5 below provides more detail of both the number of occurrences and the strength of the links among the main keywords represented in Figure 6.

Finally, Figure 7 shows the average year of publication of the keywords appearing in Figure 6.

If we take the years in which they are cited into account, we see how most of the keywords originated from the year 2010, coinciding with the period with the largest number of citations of Aitchison's text. The most recent keywords that appear in Figure 7 are "drosophila serrata," "habitat use" and "machine learning," which shows that the book continues to generate new investigations in the "geology," "ecology," and "geosciences" fields, which are the research fields that initially implemented the CoDA methodology. It is therefore to be expected that many publications based on this methodology will continue to be generated in these research fields.

Conclusions

In this paper, we have carried out a bibliometric analysis of all the publications that have cited the book entitled "The Statistical-Analysis of Compositional Data" published by John Aitchison in 1986.

We have addressed all of our established aims. With regard to the first research question (RQ1), we have analyzed how the citation structure of this work has evolved. Our analysis reveals that although the work has received citations uninterruptedly since its publication, the number of citations has increased very significantly over the past 4 years. The temporal analysis also revealed that in recent years CoDA has been mostly used in a practical way in scientific papers, although it is also true that during the period 2011 to 2019 the theoretical formulation of this methodology has started to be discussed. In reference to the major journals citing Aitchison's book (RQ2), we observe that most are indexed under the WoS categories "Geosciences, Multidisciplinary" and "Ecology." Of these, "Mathematical Geosciences" and "Computers Geosciences" stand out. The journal co-citation

Table 4. Main WoS categories that have cited Aitchison's (1986) book.

Categories of Wos	Total journals
Geosciences, Multidisciplinary	23
Ecology	19
Geochemistry & Geophysics	16
Statistics & Probability	9
Environmental Sciences	7
Evolutionary Biology	6
Geography, Physical	6
Geology	6
Multidisciplinary Sciences	6
Zoology	5
Biology	4
Mathematical & Computational Biology	4
Mathematics, Interdisciplinary Applications	4
Mineralogy	4
Entomology	3
Genetics & Heredity	3
Oceanography	3
Palaeontology	3
Soil Sciences	3
Behavioral Sciences	2
Biochemical Research Methods	2
Biochemistry & Molecular Biology	2
Biodiversity Conservation	2
Engineering, Environmental	2
Marine & Freshwater Biology	2
Microbiology	2
Water Resources	2
Automation & Control Systems	1
Biotechnology & Applied Microbiology	1
Chemistry, Analytical	1
Chemistry, Inorganic & Nuclear	1
Computer Science, Artificial Intelligence	1
Computer Science, Interdisciplinary Applications	1
Earth Sciences	1
Energy & Fuels	1
Engineering, Civil	1
Fisheries	1
Instruments & Instrumentation	1
Mining & Mineral processing	1
Physiology	1

Source. Authors based on the WoS database and the VOSviewer Software.
Note. The same journal can be indexed in two or more WoS categories.

Table 5. Most common author keywords occurrences in journals that have cited Aitchison's (1986) book.

R	Keyword	Occurrences	Total link strength
1	Compositional Data Analysis	364	174.00
2	Cuticular Hydrocarbon	109	58.00
3	Principal Component Analysis	72	48.00
4	Log-Ratio	70	54.00
5	Geochemistry	67	42.00
6	Simplex	64	48.00
7	Multivariate	57	40.00
8	Provenance	43	25.00
9	Sexual Selection	41	27.00
10	Aitchison Geometry	36	32.00
11	Isometric Log Ratio	32	23.00
12	Cluster Analysis	30	24.00
13	Dirichlet Distribution	29	18.00
14	Origin	28	27.00
15	Age	27	26.00
15	Climate	27	27.00
15	Climate-Change	27	23.00
15	Heavy-Metals	27	27.00
15	Preference	27	24.00
15	Soils	27	27.00
21	China	26	26.00
21	Profiles	26	26.00
23	Geostatistics	25	17.00
23	Habitat Selection	25	15.00
23	Constraints	25	24.00
23	Elements	25	25.00
23	Geostatistics	25	24.00
23	Impact	25	25.00
23	Natural-Selection	25	25.00
23	Statistics	25	23.00
31	Bayesian	24	7.00
31	Soil	24	15.00
31	Bayesian	24	20.00
31	River	24	23.00
31	Sea	24	20.00
31	Social Insect	24	24.00
31	Systems	24	22.00
31	Uncertainty	24	21.00
39	Archaeometry	23	22.00
39	Holocene	23	20.00
39	Missing Values	23	23.00
39	Organic-Matter	23	23.00
39	Predation	23	22.00
39	Xrf	23	22.00

Source. Authors, WoS database, 1986 through November 2019.
Note. R = ranking.

network clarifies the distribution of core journals. With regard to the third research question (RQ3), our results show how the keywords in the main papers to have cited the book correspond to the year 2010, coinciding with the period that had the greatest number of citations.

Our study presents several findings that allow us to understand the evolution and advances that are taking place in the CoDA field through an analysis of the citations received by Aitchison's (1986) book, a seminal text in both the field of geoscience and modern science. Firstly, the present work paints a collective picture of the academic structure cited in the book. Secondly, there has been a significant increase in

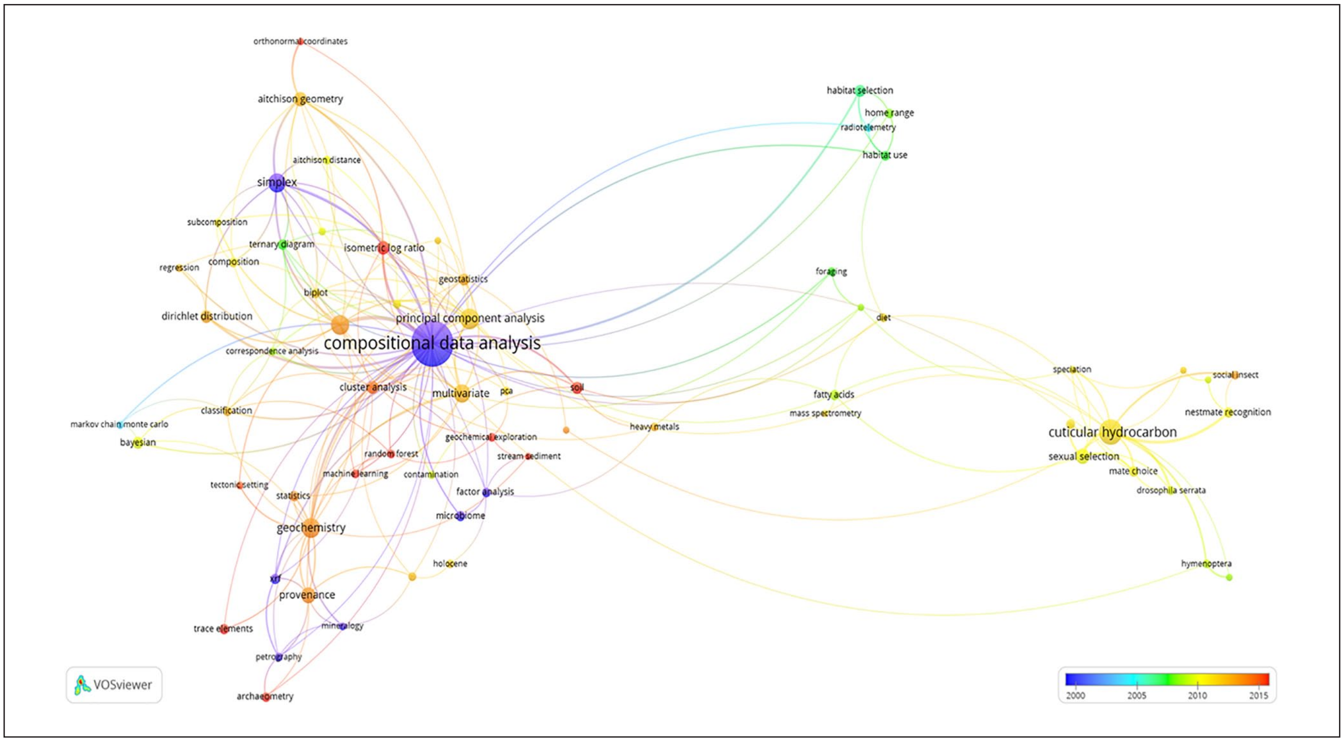

Figure 7. Co-occurrence of author keywords who have cited Aitchison's (1986) book. Citation threshold of 10 and showing the 200 most representative co-citation connections.
Note. Authors, based on WoS database; figure created using VOSviewer Software.

the number of articles citing it since 2010. In analyzing these publications, we have confirmed a high level of collaboration between different research fields, which has allowed application of the CoDA methodology to spread significantly through different academic fields. Thirdly, our analysis reveals important relationships between the main journals to cite the book and others indexed in the most prestigious quartiles in the fields related to "Geochemistry & Geophysics," "Chemistry," "Multidisciplinary Sciences," "Ecology," "Statistics & Probability," "Mathematics," "Environmental Sciences," "Zoology," but especially in the field of "geoscience," thus demonstrating the multidisciplinary nature of research using CoDA.

The co-occurrence of keyword analysis has identified research topics that have not yet been widely developed, as well as research trends that will prove useful as the amount of literature under analysis increases (Law et al., 2019). Lamberton and Stephen (2016) stated that the periodical review of the state of research makes it possible to map out the next stages of research in an innovative, relevant and rigorous way. In this regard, we believe the information contained in this paper will prove very useful for academics, since it provides them with a snapshot of the directions of foreseeable research in CoDA in the various fields that make most use of this methodology. This paper will also report on the most researched domains within the CoDA framework, thus enabling researchers to identify research gaps that will need to be filled by further studies in the future (Faruk et al., 2021). The theoretical implications of this study provide an overview of the literature on the development of CoDA research worldwide through a bibliometric analysis. This allows identification of the components of the main concept. In addition, the level of growth of the research conducted historically, the concurrence of the keywords across clusters, the leading journals and scientific collaboration on this topic of study are determined. On the other hand, the main practical implications of this study fall directly on CoDA researchers, teachers, and students. Furthermore, the methodology used in this study can be used to obtain similar results in other contexts (Quintero-Quintero et al., 2021).

The methodology used in this study presents several advantages. First, the paper shows a network map of the related journals. This allows for a more convenient tracing of the initial theoretical roots and historical context of the field. Furthermore, the keyword analysis approach adopted through temporal evolution allows researchers to follow the development of CoDA in the literature, providing opportunities to expand the current body of research (Liu et al., 2022) enabling the generation of new approaches, as well as the identification of future research trends. In summary, we have identified the main research fields and topics cited in Aitchison's (1986) book, these being addressed through a multidisciplinary and interdisciplinary approach (Altinay & Taheri, 2019).

Our document contributes to the body of relevant literature by systematizing the CoDA literature through the application of VOSviewer software as a visualized analytical tool for bibliometric analysis, providing valuable references for researchers wishing to delve deeper into this area of knowledge. In addition, it reflects CoDA's network maps and information tables in a more comprehensive way, providing a clear orientation to follow the development and then recognize emerging trends. Thirdly, it shows the most influential journals in the discipline, allowing researchers to perform precise journal searches. Finally, it can also guide scholars on how to approach a study involving knowledge mapping with the applicable analytical element of publications (Liu et al., 2022).

This paper has some limitations. The first concerns the database used to carry out the study, that is, the WoS database. For example, the WoS does not include all academic journals, and therefore journals included in other databases, such as those included in the "Emerging sources citation index," have not been considered. Another limitation of this database is that it uses a "full counting" method to collect data. In order to resolve this limitation, our research also incorporated the "fractional counting" method, using the VOSviewer software to detect the co-occurrence of author keywords and co-citations of journals. The third limitation is that not all documents indexed in the WoS were considered: only those subject to a strict process of arbitration. A further limitation is that the results are dynamic and will inevitably change over time. Despite these limitations of our analysis, we consider that this paper can be regarded as an overview of the relationships that occur between Aitchison's book and the geoscience field, as well as modern science, expanding on what is already known about the beginnings of CoDA in journals (Navarro et al., 2021).

This document also provides a starting point for future studies, as our results can be complemented by those obtained in other journals that choose to include journals appearing in the "Emerging sources citation index," as these journals offer less experienced researchers a good opportunity to publish their results. This can also lead to the development of emerging themes and new research trends (Mulet-Forteza et al., 2019) in the field of CoDA through the comparing of results, which, in turn, can lead to the development of new conceptual frameworks (Mulet-Forteza et al., 2021). It may also prove interesting to repeat this work using several databases, and not only the WoS, in order to compare the results obtained from the use of several databases. In addition, such research could highlight some of the limitations of the different databases, some of which are discussed in this paper. Finally, it is also proposed that a literary review be conducted of the main documents that cite Aitchison's (1986) book in the main scientific fields highlighted in this paper, as this will provide very relevant information on how scientific research in CoDA may progress in these fields.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Spanish Ministry of Science, Innovation and Universities/FEDER (grant number RTI2018-095518-B-C21).

ORCID iD

Carles Mulet-Forteza https://orcid.org/0000-0002-3904-8314

References

Ahmad, S., Javed, Y., Hussain Khahro, S., & Shahid, A. (2020). Research contribution of the oldest seat of higher learning in Pakistan: A bibliometric analysis of University of the Punjab. *Publications*, *8*(3), 43.

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *44*(2), 139–160.

Aitchison, J. (1986). *The statistical analysis of compositional data. Monographs on statistics and applied probability*. Chapman and Hall.

Altinay, L., & Taheri, B. (2019). Emerging themes and theories in the sharing economy: A critical note for hospitality and tourism. *International Journal of Contemporary Hospitality Management*, *31*(1), 180–193.

Batista-Foguet, J. M., Ferrer-Rosell, B., Serlavós, R., Coenders, G., & Boyatzis, R. E. (2015). An alternative approach to analyze ipsative data. Revisiting experiential learning theory. *Frontiers in Psychology*, *6*, 1742.

Belles-Sampera, J., Guillen, M., & Santolino, M. (2016). Compositional methods applied to capital allocation problems. *The Journal of Risk*, *19*(2), 15–30.

Biemans, W., Griffin, A., & Moenaert, R. (2010). In search of the classics: A study of the impact of JPIM papers from 1984 to 2003. *Journal of Product Innovation Management*, *27*(4), 461–484.

Blasco-Duatis, M., Sáez Zafra, M., & García Fernández, N. (2018). Compositional representation (coda) of the agenda-setting of the political opinion makers in the main Spanish media groups in the 2015 general election. *Communication & Society*, *31*(2), 1–24.

Bornmann, L. (2014). How are excellent (highly cited) papers defined in bibliometrics? A quantitative analysis of the literature. *Research Evaluation*, *23*, 166–173.

Bornmann, L., & Leydesdorff, L. (2014). Scientometrics in a changing research landscape. *Science and Society*, *15*, 1228–1232.

Boyack, K. W., & Klavans, R. (2014). Creation of a highly detailed, dynamic, global model and map of science. *Journal of the Association for Information Science and Technology*, *65*(4), 670–685.

Brueton, V. C., Vale, C. L., Choodari-Oskooei, B., Jinks, R., & Tierney, J. F. (2014). Measuring the impact of methodological research: A framework and methods to identify evidence of impact. *Trials*, *15*, 464.

Bruns, H. J., Christensen, M., & Pilkington, A. (2020). Intellectual heritages of post-1990 public sector accounting research: An exploration. *Accounting Auditing & Accountability Journal*, *33*(8), 2077–2110.

Buccianti, A., Lima, A., Albanese, S., Cannatelli, C., Esposito, R., & De Vivo, B. (2015). Exploring topsoil geochemistry from the CoDA (compositional data analysis) perspective: The multi-element data archive of the Campania Region (Southern Italy). *Journal of Geochemical Exploration*, *159*, 302–316.

Butt, N. S., Malik, A. A., & Shahbaz, M. Q. (2021). Bibliometric analysis of statistics journals indexed in web of science under emerging source citation index. *Sage Open*, *11*(1), 1–8.

Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, *22*(2), 191–235.

Carreras Simó, M., & Coenders, G. (2020). Principal component analysis of financial statements. A compositional approach. *Revista de Metodos Cuantitativos para la Economía y la Empresa*, *29*, 18–37.

Carreras-Simó, M., & Coenders, G. (2021). The relationship between asset and capital structure: A compositional approach with panel vector autoregressive models. *Quantitative Finance and Economics*, *5*(4), 571–590.

Chakraborty, S., Majumdar, A. S., & Shukla, A. D. (2020). Role of fluid in strain softening within the main central thrust in Sikkim: The origin of quartz-rich mylonites. *Journal of Structural Geology*, *140*, 104145.

Chen, G., Ju, B., Fang, H., Chen, Y., Yu, N., & Wan, Y. (2020). Air bearing: Academic insights and trend analysis. *The International Journal of Advanced Manufacturing Technology*, *106*, 1191–1202.

Clark, T., Key, T. M., Hodis, M., & Rajaratnam, D. (2014). The intellectual ecology of mainstream marketing research: An inquiry into the place of marketing in the family of business disciplines. *Journal of the Academy of Marketing Science*, *42*(3), 223–241.

Coenders, G., & Ferrer-Rosell, B. (2020). Compositional data analysis in tourism. Review and future directions. *Tourism Analysis*, *25*(1), 153–168.

Cook, D. J., Mulrow, C. D., & Haynes, R. B. (1997). Systematic reviews: Synthesis of best evidence for clinical decisions. *Annals of Internal Medicine*, *126*(5), 376–380.

Culnan, M. J. (1986). The intellectual development of management information systems, 1972–1982: A cocitation analysis. *Management Science*, *32*(2), 156–172.

Daunis-i-Estadella, J., Thió-Henestrosa, S., & Mateu-Figueras, G. (2011). Including supplementary elements in a compositional biplot. *Computational Geosciences*, *37*(5), 696–701.

Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing & Management*, *37*(6), 817–842.

Ding, Y., Rousseau, R., & Wolfram, D. (2014). *Measuring scholarly impact: Methods and practice*. Springer.

Diodato, V. (1994). *Dictionary of bibliometric*. Haworth Press.

Ezbakhe, F., & Pérez Foguet, A. (2020). Child mortality levels and trends: A new compositional approach. *Demographic Research*, *43*, 1263–1296.

Fan, J., Gao, Y., Zhao, N., Dai, R., Zhang, H., Feng, X., Shi, G., Tian, J., Chen, C., Hambly, B. D., & Bao, S. (2020). Bibliometric analysis on COVID-19: A comparison of research between English and Chinese studies. *Public Health Frontier*, *8*, 477. https://doi.org/10.3389/fpubh.2020.00477

Faruk, M., Rahman, M., & Hasan, S. (2021). How digital marketing evolved over time: A bibliometric analysis on scopus database. *Heliyon*, *7*(12), e08603.

Feng, G. C. (2020). Research performance evaluation in China: A big data analysis. *Sage Open*, *10*(1), 1–11.

Ferrer-Rosell, B., Coenders, G., & Martínez-Garcia, E. (2015). Determinants in tourist expenditure composition—The role of airline types. *Tourism Economics*, *21*(1), 9–32.

Filzmoser, P., & Hron, K. (2009). Correlation analysis for compositional data. *Mathematical Geosciences*, *41*(8), 905–919.

Filzmoser, P., Hron, K., & Templ, M. (2018). *Applied compositional data analysis with worked examples in R*. Springer.

García-Lillo, F., Úbeda-García, M., & Marco-Lajara, B. (2016). The intellectual structure of research in hospitality management: A literature review using bibliometric methods of the journal International Journal of Hospitality Management. *International Journal of Hospitality Management*, *52*, 121–130.

Garousi, V., & Fernandes, J. M. (2017). Quantity versus impact of software engineering papers: A quantitative study. *Scientometrics*, *112*, 963–1006.

Gatrell, C., & Breslin, D. (2017). Editors' statement. *International Journal of Management Reviews*, *19*, 1–3.

Ghorbani, Z., Kargaran, S., Saberi, A., Haghighinasab, M., Jamali, S. M., & Ale Ebrahim, N. (2021). Trends and patterns in digital marketing research: Bibliometric analysis. *Journal of Marketing Analytics*. Advance online publication. https://doi.org/10.1057/s41270-021-00116-9

Gläser, J., & Laudel, G. A. (2015). Bibliometric reconstruction of research trails for qualitative investigations of scientific innovations. *Historical Social Research*, *40*, 299–330.

Godin, B. (2006). On the origins of bibliometrics. *Scientometrics*, *68*(1), 109–133.

Graffelman, J., Pawlowsky-Glahn, V., Egozcue, J. J., & Buccianti, A. (2018). Exploration of geochemical data with compositional canonical biplots. *Journal of Geochemical Exploration*, *194*, 120–133.

Greenacre, M. (2018). *Compositional data analysis in practice*. Chapman and Hall/CRC press.

Guerreiro, C., Cachão, M., Pawlowsky-Glahn, V., Oliveira, A., & Rodrigues, A. (2015). Compositional Data Analysis (CoDA) as a tool to study the (paleo)ecology of coccolithophores from coastal-neritic settings off central Portugal. *Sedimentary Geology*, *319*, 134–146.

Gumpenberger, C., Wieland, M., & Gorraiz, J. (2012). Bibliometric practices and activities at the University of Vienna. *Library Management*, *33*, 174–183.

Haustein, S., & Peters, I. (2020). Commemorating Judit Bar-Ilan from bibliometric and altmetric perspectives. *Scientometrics*, *123*(3), 1211–1224.

He, L., Fang, H., Chen, C., Wu, Y., Wang, Y., Ge, H., Wang, L., Wan, Y., & He, H. (2020). Metastatic castration-resistant prostate cancer: Academic insights and perspectives through bibliometric analysis. *Medicine*, *99*, e19760.

He, L., Fang, H., Wang, X., Wang, Y., Ge, H., Li, C., Chen, C., Wan, Y., & He, H. (2020). The 100 most-cited articles in urological surgery: A bibliometric analysis. *International Journal of Surgery*, *75*, 74–79.

Hoque, M. A., Rasiah, R., Furuoka, F., & Kumar, S. (2021). Technology adoption in the apparel industry: Insight from

literature review and research directions. *Research Journal of Textile and Apparel*, *25*, 292–307.
Hota, P. K., Subramanian, B., & Narayanamurthy, G. (2020). Mapping the intellectual structure of social entrepreneurship research: A citation/co-citation analysis. *Journal of Business Ethics*, *166*, 89–114.
Hou, J., Yang, X., & Chen, C. (2018). Emerging trends and new developments in information science: A document co-citation analysis (2009–2016). *Scientometrics*, *115*(2), 869–892.
Hron, K., Coenders, G., Filzmoser, P., Palarea-Albaladejo, J., Faměra, M., & Matys Grygar, T. (2021). Analysing pairwise logratios revisited. *Mathematical Geosciences*, *53*, 1643–1666. https://doi.org/10.1007/s11004-021-09938-w
Huang, Y., Ding, X. H., Liu, R., He, Y., & Wu, S. (2019). Reviewing the domain of technology and innovation management: A visualizing bibliometric analysis. *Sage Open*, *9*(2), 1–16.
Jankó, F., Papp Vancsó, J., & Móricz, N. (2017). Is climate change controversy good for science? IPCC and contrarian reports in the light of bibliometrics. *Scientometrics*, *112*, 1745–1759.
Jappe, A. (2020). Professional standards in bibliometric research evaluation? A meta-evaluation of European assessment practice 2005-2019. *PLoS One*, *15*(4), e0231735.
Jiang, J., & Fan, L. (2022). Visualizing the knowledge domain of language experience: A bibliometric analysis. *Sage Open*, *12*(1), https://doi.org/10.1177/21582440211067252
Karakus, M., Usak, M., & Ersozlu, A. (2021). Emotions in learning, teaching, and leadership: A bibliometric review of Asian literature (1990–2018). *Sage Open*, *11*(1), 1–18.
Kogovšek, T., Coenders, G., & Hlebec, V. (2013). Predictors and outcomes of social network compositions. A compositional structural equation modeling approach. *Social Networks*, *35*(1), 1–10.
Köseoglu, M. A., Mehraliyev, F., Aladag, O. F., & King, B. (2021). Origins, evolution and themes of scholarly hospitality sources: 1960–2019. *International Journal of Hospitality Management*, *94*, 102817.
Koseoglu, M. A., Mehraliyev, F., & Xiao, H. (2019). Intellectual connections in tourism studies. *Annals of Tourism Research*, *79*, 1–15.
Köseoglu, M. A., Okumus, F., Dogan, I. C., & Law, R. (2019). Intellectual structure of strategic management research in the hospitality management field: A co-citation analysis. *International Journal of Hospitality Management*, *78*, 234–250.
Kunisch, S., Menz, M., Bartunek, J. M., Cardinal, L. B., & Denyer, D. (2018). Feature topic at organizational research methods: How to conduct rigorous and impactful literature reviews? *Organizational Research Methods*, *21*(3), 519–523.
Lamberton, C., & Stephen, A. T. (2016). A thematic exploration of digital, social media, and mobile marketing: Research evolution from 2000 to 2015 and an agenda for future inquiry. *Journal of Marketing*, *80*(6), 146–172.
Lampe, J., Kraft, P. S., & Bausch, A. (2019). Mapping the field of research on entrepreneurial organizations (1937–2016): A bibliometric analysis and research agenda. *Entrepreneurship Theory and Practice*, *44*, 784–816.
Law, R., Leung, D., & Chan, I. C. C. (2019). Progression and development of information and communication technology research in hospitality and tourism: A state-of-the-art review. *International Journal of Contemporary Hospitality Management*, *32*(2), 511–534.
Leung, X. Y., Sun, J., & Bai, B. (2017). Bibliometrics of social media research: A co-citation and co-word analysis. *International Journal of Hospitality Management*, *66*, 35–45.
Linares-Mustarós, S., Coenders, G., & Vives-Mestres, M. (2018). Financial performance and distress profiles. From classification according to financial ratios to compositional classification. *Advances in Accounting*, *40*, 1–10.
Liu, D., Che, S., & Zhu, W. (2022). Visualizing the knowledge domain of academic mobility research from 2010 to 2020: A bibliometric analysis using CiteSpace. *Sage Open*, *12*(1), https://doi.org/10.1177/21582440211068510
Liu, T., Hu, H., Ding, X., Yuan, H., Jin, C., Nai, J., Liu, Y., Wang, Y., Wan, Y., & Tao, X. (2020). 12 years roadmap of the sulfur cathode for lithium sulfur batteries (2009–2020). *Energy Storage Materials*, *30*, 346–366.
Lopes, M. A., & Martins, R. A. (2021). Mapping the impacts of Industry 4.0 on performance measurement systems. *IEEE Latin America Transactions*, *19*(11), 1912–1923.
McBurney, M. K., & Novak, P. L. (2002). *What is bibliometric and why should you care?* [Conference session]. Proceedings of the Professional Communication Conference, Portland, OR (pp. 108–114).
McCain, K. W. (1986). Cocited author mapping as a valid representation of intellectual structure. *Journal of the American Society for Information Science*, *37*(3), 111–122.
McCain, K. W. (1990). Mapping authors in intellectual space: A technical overview. *Journal of the American Society for Information Science*, *41*(6), 433–443.
McCain, K. W. (1991). Mapping economics through the journal literature: An experiment in journal co-citation analysis. *Journal of the American Society for Information Science*, *42*(4), 290–296.
Merigó, J. M., Cancino, C. A., Coronado, F., & Urbano, D. (2016). Academic research in innovation: A country analysis. *Scientometrics*, *108*(2), 559–593.
Merigó, J. M., Gil-Lafuente, A. M., & Yager, R. R. (2015). An overview of fuzzy research with bibliometric indicators. *Applied Soft Computing*, *27*, 420–433.
Merigó, J. M., Mulet-Forteza, C., Valencia, C., & Lew, A. A. (2019). Twenty years of tourism geographies: A bibliometric overview. *Tourism Geographies*, *21*(5), 881–910.
Merigó, J. M., & Yang, J. B. (2017). Accounting research: A bibliometric analysis. *Australian Accounting Review*, *27*(1), 71–100.
Miesch, A. T. (1969). The constant sum problem in geochemistry. In D. F. Merriam (Ed.), *Computer applications in the earth sciences* (pp. 161–176). Springer.
Mikšová, D., Filzmoser, P., & Middleton, M. (2020). Imputation of values above an upper detection limit in compositional data. *Computational Geosciences*, *136*, 104383.
Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *El Profesional de la Informacion*, *29*(1), 1–20.
Mulet-Forteza, C., Genovart-Balaguer, J., Mauleon-Mendez, E., & Merigó, J. M. (2019). A bibliometric research in the tourism, leisure and hospitality fields. *Journal of Business Research*, *101*, 819–827.
Mulet-Forteza, C., Lunn, E., Merigó, J. M., & Horrach, P. (2021). Research progress in tourism, leisure and hospitality in Europe (1969–2018). *International Journal of Contemporary*

Hospitality Management, *33*, 48–74. https://doi.org/10.1108/IJCHM-06-2020-0521

Mulet-Forteza, C., Socias Salvá, A., Monserrat, S., & Amores, A. (2020). 80th anniversary of pure and applied geophysics: A bibliometric overview. *Pure and Applied Geophysics*, *177*, 531–570.

Muller, I., Hron, K., Fiserova, E., Smahaj, J., Cakirpaloglu, P., & Vancakova, J. (2018). Interpretation of compositional regression with application to time budget analysis. *Austrian Journal of Statistics*, *47*(2), 3–19.

Navarro, C., Gonzalez-Morcillo, S., Mulet-Forteza, C., & Linares-Mustaros, S. (2021). A bibliometric analysis of the 35th anniversary of the paper” the statistical analysis of compositional data” by John Aitchison (1982). *Austrian Journal of Statistics*, *50*(2), 38–55.

Noyons, E. C. M., Moed, H. F., & Van Raan, A. F. J. (1999). Integrating research performance analysis and science mapping. *Scientometrics*, *46*(3), 591–604.

Ortells, R., Egozcue, J. J., Ortego, M. I., & Garola, A. (2016). Relationship between popularity of key words in the Google browser and the evolution of worldwide financial indices. In J. A. Martín-Fernández & S. Thió-Henestrosa (Eds.), *Compositional data analysis. Springer proceedings in mathematics & statistics* (p. 187). Springer.

Pawlowsky-Glahn, V., & Egozcue, J. J. (2020). Compositional data in geostatistics: A log-ratio based framework to analyze regionalized compositions. *Mathematical Geosciences*, *52*, 1067–1084.

Pawlowsky-Glahn, V., Egozcue, J. J., & Tolosana-Delgado, R. (2015). *Modeling and analysis of compositional data* (p. 2015). Wiley.

Pearson, K. (1897). Mathematical contributions to the theory of evolution. On a form of spurious correlation which may Arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London*, *60*(359–367), 489–498.

Perianes-Rodriguez, A., Waltman, L., & van Eck, N. J. (2016). Constructing bibliometric networks: A comparison between full and fractional counting. *Journal of Informetrics*, *10*(4), 1178–1195.

Pospiech, S., Tolosana-Delgado, R., & van Den Boogaart, K. G. (2021). Discriminant analysis for compositional data incorporating cell-wise uncertainties. *Mathematical Geosciences*, *53*, 1–20. https://doi.org/10.1007/s11004-020-09878-x

Post, C., Sarala, R., Gatrell, C., & Prescott, J. E. (2020). Advancing theory with review articles. *Journal of Management Studies*, *57*(2), 351–376.

Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation*, *25*, 348–349.

Quintero-Quintero, W., Blanco-Ariza, A. B., & Garzón-Castrillón, M. A. (2021). Intellectual capital: A review and bibliometric analysis. *Publications*, *9*(4), 46.

Ramos-Rodríguez, A. R., & Ruíz-Navarro, J. (2004). Changes in the intellectual structure of strategic management research: A bibliometric study of the Strategic Management Journal, 1980–2000. *Strategic Management Journal*, *25*(10), 981–1004.

Rodrigues, L. A., Daunís-l-Estadella, J., Mateu-Flgueras, G., & Thió-Henestrosaê, S. (2011). Flying in compositional morphospaces: Evolution of limb proportions in flying vertebrates. In V. Pawlowsky-Glahn & A. Buccianti (Eds.), *Compositional data analysis: Theory and applications* (pp. 235–254) John Wiley & Sons, Ltd.

Sajovic, I., & Boh Podgornik, B. (2022). Bibliometric analysis of visualizations in computer graphics: A study. *Sage Open*, *12*(1), https://doi.org/10.1177/21582440211071105

Salisu, S. A., & Salami, M. O. (2020). A bibliometric analysis of Nigeria’s research performance, 1901-2016. *African Journal of Library Archives and Information Science*, *30*(1), 23–36.

Sanz-Sanz, E., Martinetti, D., & Napoléone, C. (2018). Operational modelling of peri-urban farmland for public action in Mediterranean context. *Land Use Policy*, *75*, 757–771.

Sarin, S., Haon, C., & Belkhouja, M. (2018). A bibliometric analysis of the knowledge exchange patterns between major technology and innovation management journals (1999-2013). *Journal of Product Innovation Management*, *35*(1), 2–8.

Segura-Robles, A., Parra-González, M. E., & Gallardo-Vigil, M. A. (2020). Bibliometric and collaborative network analysis on active methodologies in education. *Journal of New Approaches in Educational Research*, *9*(2), 259–274.

Shafique, M. (2013). Thinking inside the box? Intellectual structure of the knowledge base of innovation research (1988-2008). *Strategic Management Journal*, *34*(1), 62–93.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*, 265–269.

Smits, P., & Champagne, F. (2020). Governance of health research funding institutions: An integrated conceptual framework and actionable functions of governance. *Health Research Policy and Systems*, *18*(1), 22.

Su, X., Li, X., & Kang, Y. (2019). A bibliometric analysis of research on intangible cultural heritage using CiteSpace. *Sage Open*, *9*(2), 1–18.

Svensson, G. (2010). SSCI and its impact factors: A ‘prisoner’s dilemma’? *European Journal of Marketing*, *44*(1–2), 23–33.

Szomszor, M., Adams, J., Fry, R., Gebert, C., Pendlebury, D. A., Potter, R. W. K., & Rogers, G. (2021). Interpreting bibliometric data. *Frontiers in Research Metrics and Analytics*, *5*, 628703.

Thomas, C. W., & Aitchison, J. (2005). Compositional data analysis of geological variability and process: A case study. *Mathematical Geology*, *37*(7), 753–772.

Thomas, M. F. (2016). New keywords in the geosciences - some conceptual and scientific issues. *Revista do Instituto Geologico*, *37*(1), 1–12.

Tolosana-Delgado, R., Mueller, U., & van Den Boogaart, K. G. (2019). Geostatistics for compositional data: An overview. *Mathematical Geosciences*, *51*(4), 485–526.

Torma, G., & Thøgersen, J. (2021). A systematic literature review on meta sustainability labeling – What do we (not) know? *Journal of Cleaner Production*, *293*, 126194.

Torraco, R. J. (2016). Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review*, *15*(4), 404–428.

Van Den Boogaart, K. G., & Tolosana-Delgado, R. (2013). *Analyzing compositional data with R*. Springer.

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538.

Van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact* (pp. 285–320). Springer.

Vasilyeva, E., Kozlov, A., Alfaro-Bittner, K., Musatov, D., Raigorodskii, A. M., Perc, M., & Boccaletti, S. (2021). Multilayer representation of collaboration networks with higher-order interactions. *Scientific Reports*, *11*, 5666.

Verma, S. P., Guevara, M., & Agrawal, S. (2006). Discriminating four tectonic settings: Five new geochemical diagrams for basic and ultrabasic volcanic rocks based on log—Ratio transformation of major-element data. *Journal of Earth System Science*, *115*(5), 485–528.

Vogel, R. (2014). What happened to the public organization? A bibliometric analysis of public administration and organization studies. *The American Review of Public Administration*, *44*, 383–408.

Wang, F., Humblé, P., & Chen, W. (2019). A bibliometric analysis of translation studies: A case study on Chinese canon the journey to the West. *Sage Open*, *9*(4), 1–13.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*, 13–23.

Ye, Q., Li, T., & Law, R. (2013). A coauthorship network analysis of tourism and hospitality research collaboration. *Journal of Hospitality & Tourism Research*, *37*(1), 51–76.

Zhang, D., Fu, H.-Z., & Ho, Y.-S. (2017). Characteristics and trends on global environmental monitoring research: A bibliometric analysis based on Science Citation Index Expanded. *Environmental Science and Pollution Research International*, *24*, 26079–26091.

Zhang, X., Estoque, R. C., Xie, H., Murayama, Y., & Ranagalage, M. (2019). Bibliometric analysis of highly cited articles on ecosystem services. *PLoS One*, *14*, e0210707.

Zhao, H., Zhang, F., & Kwon, J. (2018). Corporate social responsibility research in international business journals: An author co-citation analysis. *International Business Review*, *27*(2), 389–400.

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, *18*(3), 429–472.

Zyoud, S. H., & Fuchs-Hanusch, D. (2020). Mapping of climate change research in the Arab world: A bibliometric analysis. *Environmental Science and Pollution Research*, *27*(3), 3523–3540.

9. Article #3

Research progress in compositional data in social sciences. A bibliometric analysis

Abstract

Since the publication of John Aitchison's seminal 1982 paper 'The Statistical Analysis of Compositional Data', the number of studies that apply compositional data techniques has increased exponentially, especially in the social sciences. Therefore, the purpose of this paper is to provide a holistic description of the evolution of the use of this methodology in the social sciences from 1982 to the present. The methodology used is based on a bibliometric study of the performance of the publications analysed, as well as a scientific mapping of these publications. The performance analysis involves a whole sequence of bibliometric statistics, including the analysis of the citation structure, as well as the most influential and productive authors, institutions, and countries. The scientific mapping was carried out using VOSViewer software and consisted of a joint citation analysis of the most relevant and influential authors as well as a bibliographic linkage of the most influential and representative institutions and countries in the application of compositional data applied in the field of social sciences. The main results of the work reveal the most influential and productive authors, institutions, and countries in the application of compositional data in the field of social sciences. The findings are useful for potential authors to have a quick snapshot of what to expect from this field of research in terms of what is happening.

Keywords: Bibliometric analysis; compositional data (CoDa); research output; science mapping; WoS; VOSviewer Software.

1. Introduction

The term *Compositional Data analysis* (CoDa) has historically been defined as involving random vectors with strictly positive components whose sum is constant. More recently, the term encompasses all those vectors that represent parts of a whole and carry only relative information, thus including not only unit parts or percentages but also molar compositions.

The CoDa methodology has its origins in a paper published in 1982 by John Aitchison entitled 'The Statistical Analysis of Compositional Data'. In this paper, details a whole set of techniques based on compositional data (Daunis-i-Estadella et al. 2011; Graffelman et al. 2018; Thomas and Aitchison 2005; Tolosana-Delgado et al. 2019; Verma et al. 2006). At present, and despite the proliferation of other papers reporting the analysis of compositional data (Filzmoser et al. 2018; Greenacre 2018; Pawlowsky-Glahn et al. 2015; Van Den Boogaart and Tolosana-Delgado 2013), Aitchison's paper provides a solid foundation, so his work is universally regarded as essential and seminal.

Since the publication of Aitchison's 1982 paper, compositional data have been applied in a multitude of academic disciplines. Although concern about problems related to compositional data has been kept alive mainly by researchers in the field of geosciences, we now observe how interest in this methodology is growing, in a very important way, in all areas of the social sciences. Among other fields, compositional data have been applied in fields such as geology (geochemical elements), economics (income and expenditure distribution), medicine (body composition: fat, bones, leanness), food industry (food composition: fat, sugar, etc.), chemistry (chemical composition), ecology (abundance of different species), palaeontology (foraminifera taxa), agriculture (nutrient balance ionomics), sociology (time use studies), environmental sciences (soil pollution), genetics (genotype frequency), and more (Batista-Foguet et al. l2015; Belles-Sampera et al. 2016; Blasco-Duatis et al. 2018; Carreras Simó and Coenders 2020; Carreras-Simó and Coenders 2021; Chakraborty et al. 2020; Coenders and Ferrer-Rosell 2020; Ezbakhe and Pérez Foguet 2020; Ferrer-Rosell et al. 2015; Hron et al. 2021; Kogovšek et al. 2013. Linares-Mustarós et al. 2018; Mikšová et al. 2020; Muller et al. 2018; Ortells et al. 2016; Pawlowsky-Glahn and Egozcue 2020; Pospiech et al. 2021; Rodrigues et al. 2011; Sanz-Sanz et al. 2018).

In this article, we briefly analyse the main categories in the social sciences where compositional data analyses have been applied. As we have indicated, the main category in which such analyses have been applied is Geology. In this category, the main topics covered have focused on analysing weathering conditions, tectonic provenance and setting, and sediment and tephra composition. In the Archaeology category, the publications focus on the compositional characterisation of different techniques as well as on studies of the provenance, dating, mineralogical analysis, and variability of raw materials. In the Geochemistry Geophysics category, the topics have dealt with the interpretation of mineralogical and geochemical data in the analysis of the provenance of sedimentary deposits and thus evaluate weathering processes, sediment flows, and erosion patterns as well as the effects of anthropogenic modifications in the natural environment. As for the category Environmental Sciences Ecology, the interaction among the natural environment, the biosphere, and society have been analysed, mainly in specific cases such as fire disturbance, air pollution, and recycling of plastic waste. In regard to the Engineering category, the topics deal with solving technical problems in a comprehensive way in an attempt to provide solutions to concrete problems such as recurrent convolutional networks for visual recognition, lignin biorefinery, and spatial variability of fine particle mass concerning air pollution, among others. In Anthropology, topics related to humanity, its chronology, cultures, and forms of organisation are analysed, such as the composition of ceramics, geological provenance of artefacts, and the chronological and compositional patterns of glass. The Public Environmental Occupational Health category develops topics on the intersection between the environment and public health, such as analysis of iodine concentrations in groundwater, adiposity, and the isotemporal substitution of physical activity as well as the association of arsenic metabolism with cancer, among others. In Chemistry, the main topics analysed are related to the study of matter, its composition, and its properties and how its structures are transformed after undergoing different processes or reactions that affect its molecules and atoms. To be specific, we examined cases such as biodiversity and nutrition, implications for archaeometallurgical connections, chemical characterisation, and the variability of raw materials, as well as their contribution to the final composition of materials. In Computer Science, we analysed topics related to the study of data networks and the software needed to process the information automatically, such as the evaluation of spatial similarities and land use change and harmonic polynomial GMDH learning networks for time series modelling. In Mathematics, we examined studies of the relationships between quantities, magnitudes, and properties and logical operations; in this case, we noted examples such as evaluating and helping the exploration of association patterns in gene expression data and estimation of a system of demand-sharing equations. The Food Science Technology category focuses on ensuring safety and quality by taking into account chemical, biological, and physical properties, so we focused on works that analysed the nutritional composition of quinoa, the relationship between the time of harvest and the composition of wine, or biodiversity and nutrition. Finally, in the Business Economics category, research focuses on solving business problems through the application of theories and methods of economic science, we focused on studies of the analysis of composition data and zeros in microdata and the role of governance in the composition of foreign aid flows, among other noteworthy works.

Therefore, it can be observed how compositional data analysis has expanded, since the publication of Aitchinson's seminal work in 1982, in a very important way in the main research areas of the social sciences. This is why this paper presents a bibliometric analysis of the evolution of the compositional data methodology (CoDa) applied to the Social Sciences (Castanha and Grácio 2014), with the ultimate goal of tracking and mapping the complete picture of scientific knowledge in the field.

The information we expected to gather from the analysis should answer the following research questions (RQs):

- RQ1: What is the academic structure of the application of CoDa methodology in the field of social sciences?
- RQ2: Which are the most productive authors in application of the CoDa methodology in the area of social sciences?
- RQ3: Which institutions are the most productive in terms of publishing articles using the CoDa methodology in the social sciences?
- RQ4: Which countries have published the most papers with CoDa methodology applied in the area of social sciences?

To achieve our objectives, we used the Web of Science (WoS) database and VOSviewer software (Van Eck and Waltman 2010) to graphically map the bibliographic material.

Several contributions are presented in this paper. First, bibliometric analysis has contributed to a growing body of articles summarising achievements and trends in the field of long-term social science research. Identifying the citation structure, origin, and evolution of citations helps identify intellectual linkages in academic fields applying CoDa in the social sciences (Köseoglu et al. 2019; Shafique 2013). In this sense, mapping intellectual connections contributes to the creation of new theories and the development of existing ones, thereby providing insight into future directions that scientific research may take (Köseoglu et al. 2021). Therefore, scholarly analysis of these developments can help researchers determine the potential effects that theory might have on society. In addition, these processes provide valuable information to researchers and experts (Torraco 2016), and provide them with a study that helps deepen their understanding of the current state of CoDa research; this is especially helpful to researchers who are not very familiar with this methodology (Jiang and Fan 2022). Second, this research was conducted to deepen the intellectual connections of a wide range of studies covering various fields related to the social sciences. Third, the work covers a long period of time, allowing researchers to get a complete picture of the area covered as well as its development.

The remainder of this paper is divided into the sections. In the next section, we review he literature on the use of bibliometric methods. We then describe the methodology used in this project. Next, we discuss the results we obtained, and in the last section we summarise the main conclusions and limitations and make suggestions for future lines of research.

2. Literature Review

Bibliometrics, the main objective of which is to measure the output side of science, emerged in the early 20th century, when psychologists began to collect statistics on publications related to their field of research (Godin 2006), although it was not until the 1950s when the American chemist Eugene Garfield, because of the exponential growth of academic publications, began to evaluate and make systematic counts of publications based on the literature used and cited.

Bibliometric analysis is a fundamental statistical tool for analysing the state of knowledge in a given scientific area. Indeed, bibliometrics measures, i.e., the number of documents published and the number of citations associated with these documents. Furthermore, bibliometrics allows the results of the analysis to be mapped by highlighting the connections among the main publications, authors, institutions, topics, and other characteristics of the field under study (Gumpenberger et al. 2012; Vogel 2014).

Bibliometrics is the quantification of scholarly output based on certain classifications that provide indirect indications of the perception of scholarly output. Nowadays, there are multiple definitions of bibliometrics, although the modern definition of the term is usually attributed to Alan Pritchard (1969), who defined bibliometrics as 'the application of mathematical and statistical methods to books and other means of communication'. More recently, other authors have made multiple definitions of the term. For example, Zupic and Cater (2015) indicated that bibliometrics is an instrument for analysing the evolution of disciplines based on intellectual, social, and conceptual structures, and Merigó et al. (2017) indicated that bibliometrics is a research field that quantitatively studies bibliographic material by analysing a research area and identifying its leading trends. Therefore, we can indicate that bibliometrics is the discipline that attempts to evaluate and map the progress made in scientific fields through the classification of data, including, among others, the analysis of research performance by individuals, institutions, countries and the mapping of the structure of the analysed field (Cobo et al. 2011), and so on, through statistical techniques (Diodato 1994; McBurney and Novak 2002).

One application of bibliometric methods is their use as a tool to evaluate research that has been conducted (Bornmann and Leydesdorff 2014). This is the easy part of bibliometrics because it provides straightforward information and requires no assumptions to be made. However, trying to assess the quality and importance of published papers is a much more complex and less straightforward task. At present, two main methods are used to try to assess the quality of published papers, one qualitative (peer review) and the other quantitative (bibliometrics). In this respect, the peer review method involves subjective, nonquantifiable assessments by competent experts, whereas the bibliometric method has opted for a simpler path, still in use today, that considers that a document becomes more important as the number of citations it receives increases.

Nowadays, new alternatives to the classic citation for assessing the importance of a scientific paper have emerged, such as libmetrics and altimetrics, among others. *Libmetrics* establishes a connection between the importance of a scientific article or book and its availability in a library by measuring, i.e. how often it is acquired by or borrowed from the library. *Altimetrics*, on the other hand, generates new knowledge by combining all available online data and the application of big data technologies. These alternative bibliometric methods are based on free online content that complements the data offered by bibliometrics based on conventional databases, such as the WoS or Scopus.

Today, Web of Science (WoS) and Scopus are the most widely used databases for conducting bibliometric studies. A third database, Google Scholar, was launched in 2004, although this platform has not yet become an alternative to the two previous databases due to significant shortcomings in comparison with its predecessors (Jacso 2005; Neuhaus et al. 2006; Falagas et al. 2008; Giustini and Boulos 2013; Halevi et al. 2017), such as the fact that it is not entirely clear which publications are accepted in this database or how citations of these publications are produced. This results in a lack of transparency with regard to the data offered, making it difficult to carry out reliable bibliometric analyses. In addition, the relatively low quality of the metadata available in Google Scholar and the difficulty of extracting them (Orduña-Malea et al. 2016) mean that using data from Google Scholar for bibliometric studies represents quite a challenge (Harzing 2016; Harzing and Alakangas 2016; Moed et al. 2016; Halevi et al. 2017; Martín-Martín et al. 2018; López-Cózar et al. 2019). That being said, several studies have highlighted the potential of Google Scholar as a free and more complete alternative to the WoS and Scopus databases (Aguillo 2012; Gehanno et al. 2013; Haddaway et al. 2015; Moed et al. 2016; Martín-Martín et al. 2020), especially when we consider that many comparisons of literature search databases have focused on citation analysis and coverage of bibliographic records (Meho and Yang 2007; Kousha and Thelwall 2008; Mikki 2010; Martín-Martín et al. 2018; Martín-Martín et al. 2020) and not on the effectiveness of the platforms in retrieving relevant articles through keyword searches. In fact, it is often pointed out that Google Scholar covers many more records than the other two databases.

With respect to the above, several papers have analysed the differences in record coverage and citation data between Google Scholar, Scopus and WoS. The latter covers more than 75 million records in its core collection, and up to 155 million records when other regional and subject-specific citation indexes are included (Birkle et al. 2020). Scopus claims to cover more than 76 million records (Baas et al. 2020), while, according to several recent studies, Google Scholar covers more than 400 million records (Gusenbauer 2019; López-Cózar et al. 2019). This lack of coverage by the two main databases, namely WoS and Scopus, has meant that a large part of the academic literature has been excluded from their records, especially that espousing the epistemologies and worldviews of the Global South (Corona Berkin 2017). Several studies have shown that scientific research carried out by authors from Latin America, Africa and Central European countries is not cited in the main social science journals (Beigel 2013a, 2013b; Alonso-Gamboa and Espinosa-Reyna 2015; Beigel 2016; Mosbah-Natanson and Gingras 2015), meaning a substantial amount of research is excluded from the great conversation that is the circulation of international knowledge. It is not possible to speak of universal knowledge when certain research is only produced and reproduced in a limited geographical context and focused on certain geographical areas, certain languages and certain disciplines (Beigel 2014). In contrast to the above, Google Scholar is a dynamic, open and uncontrolled database that covers all languages, typologies and disciplines (Martín-Martín et al. 2016).

The results of bibliometric studies, in addition to being of great interest to researchers themselves, provide very useful information for policymakers and academic decision makers in universities, research centres, and governments. In this respect, outstanding papers in bibliometric studies are considered reliable and relevant sources of results and are often used to justify decisions concerning research policies, job offers, and promotions, as well as to direct and support research projects (Bornmann and Leydesdorff 2014; Gläser and Laudel 2015). In addition, both public and private research funding agencies often require researchers to either provide certain indications of quality before funding their research or to demonstrate that the research to be conducted has the potential to affect society in some way (Bornmann 2014; Bornmann and Leydesdorff 2014; Brueton et al. 2014). Finally, we want to point out that bibliometrics can help journal editors evaluate past publications, design new policies, and make future editorial decisions.

3. Methodology

The data used to conduct the bibliometric study were compiled from the WoS in May 2022. According to Merigó et al. (2015), the WoS is considered the most influential database in the world.

To obtain the data, we first had to select the WoS publications that met the requirements of the study, that is, that analysed the impact of the CoDa methodology in the area of social sciences. To do this, several filters were applied in the WoS. The first was a search by topic. We selected ‘compositional data’, ‘compositional data analysis’, ‘CoDa’, ‘CoDA’, ‘CODA’, ‘log ratio’, and ‘biplot’. The search returned a total of 34,910 records. We then filtered these by the research domain ‘social sciences’ and removed all records before 1981 because the first publication in ‘Compositional Analysis’ was by John Aitchison in 1982 (Navarro-Lopez et al. 2021). After we applied these filters, 7,802 records remained. we then eliminated certain research areas because they do not use compositional data.

The following research areas were eliminated: acoustics, audiology, speech–language pathology, music, respiratory system, philosophy, literature, theatre, otorhinolaryngology, history philosophy of science, religion, art, ophthalmology, and fisheries. After we eliminated these research areas, the number of records was reduced to 5,501. Finally, we reviewed these 5,501 papers one by one to select those that dealt exclusively with applications of compositional data in the social sciences. This review reduced the number of records we analysed to 1,155.

Various bibliometric techniques have been applied to these 1,155 records because there is no consensus in the academic literature on the best bibliometric methods to use. Quantitative methods have been applied, which in turn have been broken down into *evaluative techniques* and *relational techniques* (Benckendorff and Zehrer 2013; Cobo et al. 2011). The former assesses both the productivity and the impact of a scientific paper. These include the number of documents, which assesses productivity (Ding et al. 2014); the number of citations, which measures the influence of publications (Svensson 2010); and the ratio of citations per document, which determines the average impact of each publication; and the h-index, which measures both impact and productivity by determining the number “h” of publications that have at least “h” citations. These evaluative techniques have been applied to the analysis of the most productive and influential authors, institutions, and countries. They also have been used in numerous previous bibliometric studies, such as those of Mauleon-Mendez et al. (2018, 2020), Martorell-Cunill et al. (2019), and Mulet-Forteza et al. (2019, 2020), among many others.

The related techniques used are coauthorship and bibliographic coupling. *Coauthorship* occurs when a document is produced by more than one author, and *bibliographic coupling* occurs when two documents cite the same third document (Kessler 1963). It is a measure of similarity between documents (Zupic and Cater 2015), which increases as the number of common references of those documents increases. Thus, it boosts the probability that the two documents connected by bibliographic coupling analyse a related topic (Martyn 1964). Bibliographic coupling can also be applied to institution and country analyses. Two institutions (or countries) are bibliographically coupled if they share a common reference to a third document (Zhao and Strotmann 2008). The above-mentioned relational analyses have been carried out using VOSviewer software (Van Eck and Walkman 2010) which has allowed the results to be graphically mapped for relational analysis. Such maps allow one to monitor various aspects of a scientific field (Noyons et al. 1999; Su et al. 2019), thus providing a clearer view of the results obtained (Merigó et al. 2016). This software improves on the results obtained previously by using fractional counting and allowed us to assess the effectiveness of coauthorship in the analysis (Merigó et al. 2019).

Before we could perform the graphical analysis with the VOSviewer software, we had to clean the data collected from the WoS, so we unified them to avoid the use of duplicate data; that is, names of the same author, institution, and country that appeared under different spellings were unified. For example, the author names 'Ods, T.', 'Olds Tim', and 'Olds, Timothy' were unified into a single name; in terms of institutions, the institutions belonging to The French National Centre for Scientific Research in France were unified; and, in the case of a country, England, Wales, Scotland, and Noth Ireland in the United Kingdom were unified.

The combination of methods used in this project allowed us to collect data using both full count and fractional count methods. With the former, each researcher is assigned a publication coauthored by several researchers with a total weight of 1, whereas the fractional count method (VOSviewer software) divides the authorship of the paper by the number of authors (Mulet-Forteza et al. 2019). In this sense, it should be borne in mind that the development of bibliometric networks is not a trivial process and, depending on how it is done, can yield very different results, as Perianes-Rodríguez et al. (2016) pointed out.

4. Results

We now turn to the various RQs we posed.

4.1. Academic Structure of the Application of CoDa Methodology in the Field of Social Sciences

In this section, we present the publication and citation structure of the application of the CoDa analysis methodology in the field of social sciences, thus answering RQ1.

Table 1. General Citation Structure in the Social Sciences Field According to the Web of Science Database

Year	≥ 100	≥ 50	≥ 25	≥ 10	≥ 5	≥ 1	TP	TC	H-Index
1988	0	0	1	1	1	2	2	35	1
1989	0	0	0	0	0	0	0	0	0
1990	0	0	0	0	0	0	1	0	0
1991	0	0	2	3	3	4	4	72	3
1992	0	2	2	3	3	3	3	135	3
1993	0	0	2	4	6	6	6	115	6
1994	1	1	2	2	2	2	2	160	2
1995	0	0	0	2	3	3	3	35	3
1996	1	2	3	5	5	5	5	256	5
1997	1	1	3	3	4	4	4	205	4
1998	1	1	1	1	3	4	4	123	4
1999	1	2	4	5	6	6	6	290	6
2000	2	4	4	7	7	7	7	503	7
2001	1	2	4	8	9	10	10	303	8
2002	3	5	8	9	10	11	11	729	9
2003	1	2	5	6	6	8	8	560	6
2004	1	5	5	8	8	8	8	604	8
2005	1	2	6	8	8	8	8	375	8
2006	3	8	10	10	11	12	13	1.006	10
2007	1	3	7	14	15	15	15	562	12
2008	1	2	11	19	20	23	23	716	16
2009	1	3	10	16	21	23	23	860	15
2010	2	7	14	22	24	26	27	899	16
2011	2	7	11	20	25	28	30	973	15
2012	1	7	17	29	34	38	38	1.033	20
2013	4	6	11	20	25	36	37	1.357	17
2014	0	5	13	23	32	41	41	825	16
2015	2	7	31	59	71	76	78	2.224	26
2016	3	13	25	51	69	77	80	2.149	25
2017	2	11	31	53	70	88	93	2.544	29
2018	1	9	28	74	104	129	133	2.290	26
2019	0	2	8	46	89	125	135	1.276	17
2020	0	3	7	37	71	144	155	1.160	16
2021	1	1	1	8	19	88	123	432	9
2022	0	0	0	0	0	4	19	4	1
Total	**38**	**123**	**287**	**576**	**784**	**1.064**	**1.155**	**24.810**	
Porcentaje	**3,29%**	**10,65%**	**24,85%**	**49,87%**	**67,88%**	**92,12%**	**100%**		

Notes: ≥100, ≥50, ≥25, ≥10, ≥5, ≥1= number of papers with at least 100, 50, 25, 10, 5, and 1 citation; TP= total papers; TC= total citation; H= h-index.
Source: Web of Science, June 2022.

The results of the academic structure of all papers that have used the CoDa methodology in the field of social sciences are presented in Table 1. The data indicate a significant growth in the number of publications and citations over the years. In fact, since 2007 the number of publications has not stopped increasing significantly. We can distinguish two periods. The first is from 1988 to 2006, when there was a progressive growth with certain ups and downs, and we could say that there has been an average of six publications per year. The second period, from 2007 onward, saw an exponential growth in publications, reaching an average of 66 publications per year.

To analyse the citation structure, we recommend using minimum citation thresholds, which indicate the number of publications that received citations above the threshold. This makes it possible to determine the periods in which the highest number of highly cited publications were published. The results imply that a paper is considered highly cited in these fields if it accumulates more than 100 citations (Mulet-Forteza et al. 2019). In fact, we can see that since 1991 there have been publications with 100 or more citations except for the years 1995, 2019, and 2020.

Of the total number of papers published (1,155), 92% have received at least one citation, although the number of citations received has grown exponentially since 2007.

Table 2. Categories With the Highest Number of Journals According to the Web of Science

R	CATEGORIES	TP	TC	H
1	Geology	301	5.773	39
2	Archaeology	167	2.722	33
3	Geochemistry Geophysics	150	3.882	35
4	Environmental Sciences Ecology	146	3.636	34
5	Engineering	92	2.750	26
6	Anthropology	73	1.504	24
7	Public Environmental Occupational Health	67	918	14
8	Chemistry	59	1.596	21
9	Computer Science	54	1.499	16
10	Mathematics	53	1.112	15
11	Food Science Technology	52	1.052	17
12	Business Economics	49	991	20
13	Mineralogy	47	611	15
14	Science Technology Other Topics	44	1.219	16
15	Physical Geography	38	1.219	18
16	Remote Sensing	31	589	12
17	Nutrition Dietetics	29	1.252	17
18	Imaging Science Photographic Technology	27	545	11
19	Energy Fuels	26	480	10
20	Psychology	26	332	8
21	Agriculture	25	333	11
22	Water Resources	25	346	9
23	Mining Mineral Processing	21	135	6
24	Health Care Sciences Services	20	661	10
25	Social Sciences Other Topics	19	232	9

Notes: R: Ranking. TP= total papers; TC= total citation. H= h-index.
Source: Web of Science, June 2022.

The data in Table 2 show how citations have been distributed among the different categories that make up the area of social sciences, according to the WoS. The table shows that Geology is the category that has received the most citations, has published the most articles, and has the highest "h" number, given that it is a category in which application of the CoDa methodology has been implemented for many years. In terms of total number of publications, the next positions are headed by Archaeology and Geophysical Geochemistry. These positions change slightly both in terms of the total number of citations and in the h-index, with Geophysical Geochemistry and Environmental Sciences Ecology moving into second and third position, respectively.

4.2. Most Productive and Influential Authors

In this section, we identify the 51 most productive authors and their co-citation structure, in order to answer RQ2.

Table 3. Most Productive Authors in the Area of Social Sciences According to the Web of Science.

R	AUTHORS	INSTITUTION	≥100	≥50	≥25	≥10	≥5	≥1	TP	TC	H	SC	TC/TP
1	Dumuid, D	University of South Australia (Australia)	1	6	9	16	19	27	28	735	13	62	26,3
2	Olds, T	University of South Australia (Australia)	1	6	9	14	16	19	19	682	12	35	35,9
3	Martin-fernandez, JA	University of Girona (Spain)	1	7	10	13	16	16	17	695	11	12	40,9
4	Chastin, SFM	Glasgow Caledonian University (UK)	2	2	7	11	12	15	15	879	11	27	58,6
5	Hron, K	Palacky University Olomouc (Czech Republic)	1	3	4	9	10	13	14	401	9	21	28,6
6	Glascock, MD	University of Missouri (USA)	0	1	2	6	8	10	13	159	7	1	12,2
7	Coenders, G	University of Girona (Spain)	0	0	1	9	10	13	13	156	9	27	12,0
8	Palarea-albaladejo, J	Biomathematics & Statistics Scotland (UK)	1	3	5	8	11	12	12	687	9	17	57,3
9	Buccianti, A	University of Florence (Italy)	0	0	2	6	10	12	12	167	8	10	13,9
10	Holtermann, A	Univ Southern Denmark (Denmark)	0	0	1	3	7	10	12	120	5	13	10,0
11	Ferrer-Rosell, B	Universitat de Lleida (Spain)	0	0	1	8	10	11	11	141	9	25	12,8
12	Chaput, JP	University of Ottawa (Canada)	2	5	8	8	8	9	9	645	8	10	71,7
13	Maher, C	University of South Australia (Australia)	0	3	5	8	8	9	9	295	8	8	32,8
14	Lima, A	University Napoli Federico II (Italy)	0	0	2	8	9	9	9	170	8	18	18,9
15	De Vivo, B	University of Pegaso (Italy)	0	0	2	8	8	9	9	170	8	18	18,9
16	Albanese, S	University Napoli Federico II (Italy)	0	0	2	7	7	9	9	158	7	15	17,6
17	Gupta, N	University of Gavle (Sweden)	0	0	1	3	5	7	9	107	5	12	11,9
18	Tremblay, MS	Children's Hospital of Eastern Ontario (CHEO) (Canada)	2	4	8	8	8	8	8	626	8	10	78,3
19	Neff, H	California State University, Long Beach (USA)	0	1	4	6	8	8	8	234	7	1	29,3
20	Pawlowsky-glahn, V	University of Girona (Spain)	0	0	2	3	5	8	8	106	5	1	13,3
21	Hallman, DM	University of Gavle (Sweden)	0	0	1	2	6	8	8	93	5	9	11,6
22	Mathiassen, SE	University of Gavle (Sweden)	0	0	1	2	5	7	8	89	5	13	11,1
23	Pedisic, Z	Technical University of Liberec (Czech Republic)	1	4	4	7	7	7	7	376	7	8	53,7
24	Baxter, MJ	Nottingham Trent University (UK)	0	2	4	6	7	7	7	256	7	8	36,6
25	Gaba, A	Palacky University Olomouc (Czech Republic)	0	0	0	4	4	5	7	62	4	9	8,9
26	Stefelova, N	Palacky University Olomouc (Czech Republic)	0	0	0	4	4	5	7	62	4	9	8,9
27	Rasmussen, CL	National Research Centre for the Working Environment (Denmark)	0	0	0	1	4	5	7	41	4	6	5,9
28	Lewis, LK	Palacky University Olomouc (Czech Republic)	1	4	6	6	6	6	6	415	6	5	69,2
29	Bloise, A	University of Calabria (Italy)	1	1	3	4	4	6	6	212	4	6	35,3
30	Miriello, D	Polytechnic University of Catalonia (Spain)	1	1	2	3	3	5	6	176	4	8	29,3
31	Verma, SP	National Autonomous University of Mexico (Mexico)	0	2	2	4	5	6	6	145	5	8	24,2
32	Thiombane, M	University of Naples Federico II (Italy)	0	0	1	6	6	6	6	108	6	9	18,0
33	Meinhold, G	University of Gottingen (Germany)	0	0	2	4	6	6	6	104	6	6	17,3
34	Filzmoser, P	Vienna University of Technology (Austria)	0	0	1	4	5	6	6	97	5	1	16,2
35	Li, Y	University of North Carolina at Chapel Hill (USA)	0	0	1	4	5	6	6	90	5	1	15,0
36	Owen, N	Baker Heart & Diabet Institute (Australia)	0	0	1	2	3	6	6	63	3	2	10,5
37	Egozcue, JJ	Polytechnic University of Catalonia (Spain)	0	0	1	1	2	6	6	43	4	1	7,2
38	Carson, V	University of Alberta (Canada)	1	2	5	5	5	5	5	368	5	5	73,6
39	Standage, M	University of Bath (UK)	1	3	5	5	5	5	5	356	5	4	71,2
40	Fogelholm, M	Universtiy of Helsinki (Finland)	1	3	5	5	5	5	5	356	5	4	71,2
41	Smith, VC	University of Oxford (UK)	1	3	5	5	5	5	5	337	5	6	67,4
42	Zuo, RG	China University of Geosciences, Wuhan (China)	0	1	4	5	5	5	5	208	5	3	41,6
43	Lloyd, CD	Queen's University Belfast (UK)	0	1	4	5	5	5	5	154	5	5	30,8
44	Von Eynatten, H	University of Gottingen (Germany)	0	1	3	5	5	5	5	145	5	2	29,0
45	Jorgensen, MB	University of Copenhagen (Denmark)	0	0	1	2	4	4	5	79	4	3	15,8
46	Dygryn, J	Palacky University Olomouc (Czech Republic)	0	0	0	3	3	4	5	44	4	7	8,8
47	Ferguson, JR	University of Missouri Columbia (USA)	0	0	0	2	3	5	5	31	3	0	6,2
48	Clark, CCT	Coventry University (UK)	0	0	0	2	2	4	5	26	2	5	5,2
49	Perez-foguet, A	Polytechnic University of Catalonia (Spain)	0	0	0	0	3	4	5	25	8	3	5,0
50	Duncan, MJ	Coventry University (UK)	0	0	0	2	2	3	5	24	2	3	4,8
51	Wood, JR	The UCL Institute of Archaeology (UK)	0	0	0	1	1	5	5	22	3	4	4,4

Notes: ≥100, ≥50, ≥25, ≥10, ≥5, ≥1 = number of papers with at least 100, 50, 25, 10, 5, and 1 citations. R = ranking; TP = total papers; TC = total citations; SC = self-citations; H = h-index; TC/TP = citations per paper.
Source: Web of Science, June 2022.

Dorothea Dumuis (University of South Australia) is the most productive author, with a total of 28 publications, and the second most influential, with a total of 735; she also has the highest number of h-indexes, which shows that many of her papers are highly cited. She is followed by other highly productive authors, such as Timothy Olds (University of South Australia), Jose Antonio Martin-Fernandez (Universitat de Girona, Spain), Sebastien FM Chastin (Glasgow Caledonian University, United Kingdom) and Karel Hron (Palacky University Olomouc, Czech Republic), with a total number of publications of 19, 17, 15, and 14, respectively.

In terms of the number of citations, Sebastien FM Chastin (Glasgow Caledonian University, United Kingdom) tops the list with a total of 879 citations. He is followed, in this order, by Dorothea Dumuis (University of South Australia), José Antonio Martin-Fernandez (Universitat de Girona, Spain), Javier Palarea-Albaladejo (*Biomathematics & Statistics Scotland* - BioSS[1], UK) and Timothy Olds (University of South Australia). Of particular note are the citations obtained by Jean-Philippe Chaput (University of Ottawa, Canada), Mark S. Tremblay (Children's Hospital of Eastern Ontario, Canada), and Lucy K. Lewis (Palacky University Olomouc, Czech Republic), Lewis (Palacky University Olomouc, Czech Republic), with a lower number of publications than the other authors, they occupy positions 6, 7, and 8 of the ranking in terms of number of citations. With respect to the h-index, we can observe that all the authors in Table 3 have obtained an h-index higher than 2; with respect to this index, only four (Dorothea Dumuis, Timothy Olds, José Antonio Martin-Fernandez, and Sebastien F. M. Chastin) of the 51 authors have an h-index of two digits, none of them exceeding 20. These are Dorothea Dumuis, Timothy Olds, José Antonio Martin-Fernandez, and Sebastien F. M. Chastin.

On the other hand, we observed that 16 authors have publications that have received more than 100 citations, and this figure improves when we analysed the publications that have received 50 or more citations because in this case the number rises to 24 authors. This ranking is headed by Jean-Philippe Chaput (University of Ottawa, Canada), with two publications that have received more than 100 citations and five that have received more than 50 citations; in the second position is Mark S. Tremblay (Children's Hospital of Eastern Ontario, Canada), who has two publications in the range of 100 or more citations and four publications in the range of 50 to 100 citations. It is noteworthy that these authors have a total of nine and eight publications, respectively, despite the small number, these publications have a high impact. Finally, we should point out that 36 authors manage to cite all their publications, and another 15 authors manage to cite practically all of their works (except for a maximum of two).

Another interesting fact is related to the institutions; note in Table 3 that the authors are distributed across 35 universities, with the most representative institution being the Palacky University Olomouc (Czech Republic), with a total of five authors. The ranking continues with the University of South Australia, the University of Girona (Spain), the University Napoli Federico II (Italy), the University of Gavle (Sweden), and Polytechnic University of Catalonia (Spain), with three authors each, and the University of Göttingen (Germany) and Coventry University (United Kingdom), with two authors each. Finally, only 14 countries are represented in Table 3. The United Kingdom leads the ranking, with nine authors, closely followed by Spain, with seven, and the Czech Republic and Italy, with six each.

We now turn to a graphical analysis of the coauthorship relationships established between the main authors who have used the CoDa methodology in the social sciences.

[1] *Biomathematics & Statistics Scotland* - BioSS: Although formally a part of the James Hutton Institute, BioSS interacts scientifically with external organisations as if it were independent. This perceived independence is central to the success of BioSS: it is enshrined in recognition of the James Hutton Institute acting as "guardian" of the BioSS for its key stakeholders. https://www.bioss.ac.uk/aboutBioSS.html.

Figure 1 graphically presents the results of these relationships with a minimum threshold of five documents and the top 100 connections.

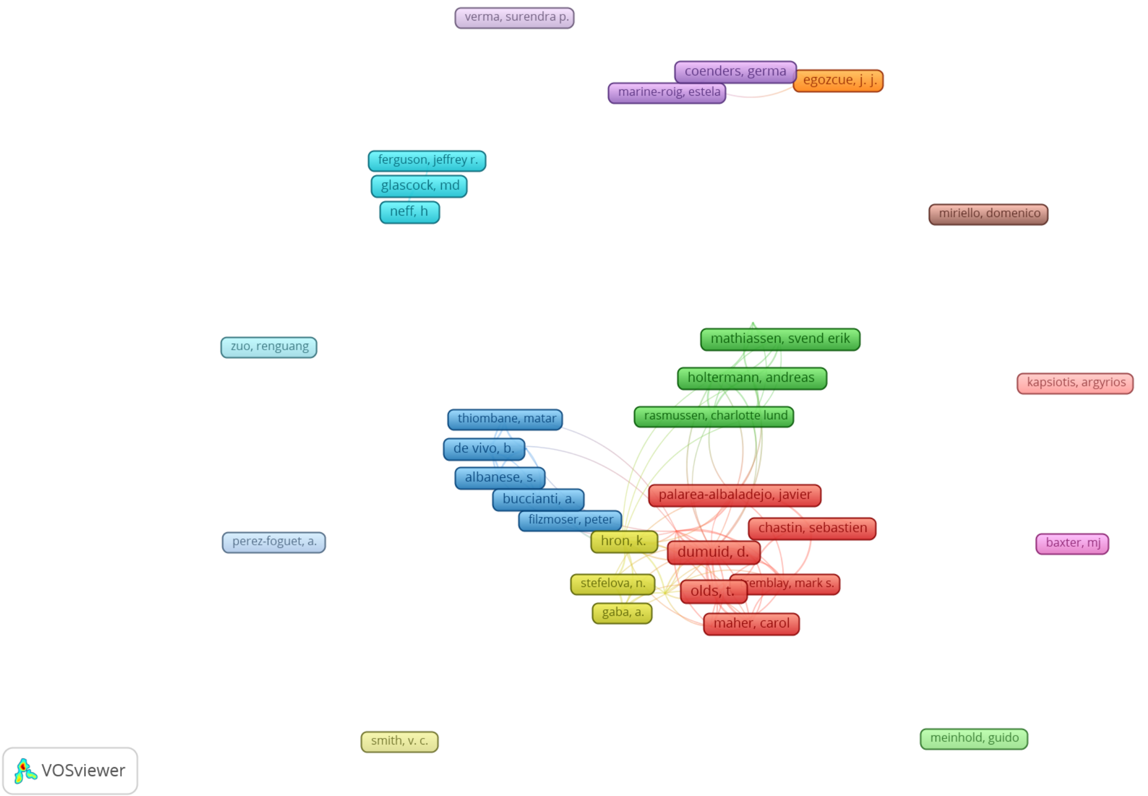

Figure 1. Coauthorship with a threshold of five documents and the 100 most representative connections
Source: VOSviewer software.

Figure 1 shows a map of co-authorship relationships between the most productive and influential authors. It shows 15 clusters of authors working together, of which eight are coauthorship clusters and seven are single-author clusters formed by a single author. The following figures show the information in Figure 1 in greater detail.

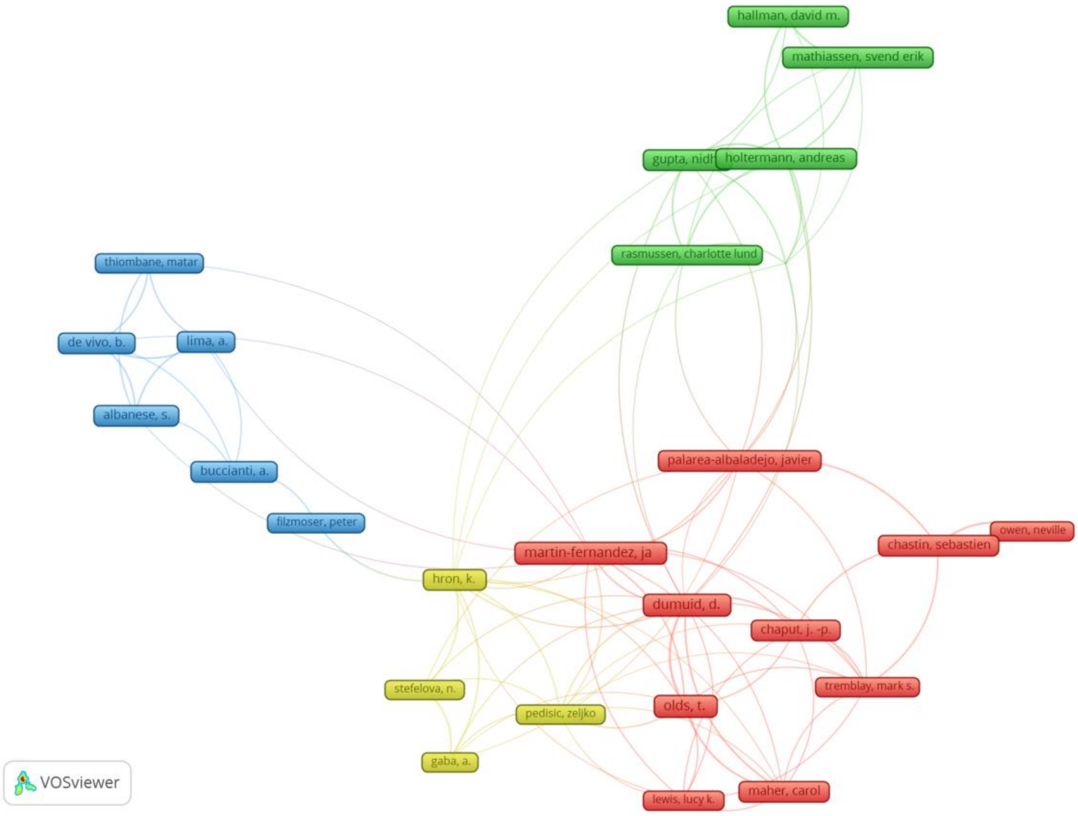

Figure 2. Coauthorship with a threshold of five papers and the 100 most representative connections
Source: VOSviewer software.

The group with the highest number of authors (in red in Figure 2) has 10 representatives, including Dorothea Dumuid, Carol Maher, Jose Antonio Martín-Fernández, Timothy Olds, Jean-Phillipe Chaput, Neville Owen, Javier Palarea-Albaladejo, and Mark S. Tremblay. As for the relationship between them, we can see in Table 3 that Dorothea Dumuid, Carol Maher, and Timothy Olds are affiliated with the University of South Australia. On the other hand, as for the relationship of the authors with the country, we can observe that to the three previous authors we can add Neville Owen, in Australia, and two authors in Canada, Mark S. Trembley and Jean-Phillipe Chaput. The most prominent author in this cluster is Dorothea Dumuis, who in turn has relationships with authors from other clusters, such as Charlotte Lund Rasmussen, Zelijko Pedisic, and Nikole Stefelova. The second largest cluster (in green in Figure 2) has six authors, centred on Sven Erik Mathiassen, Charlotte Lund Rasmussen, and Andreas Holtermann. The data in Table 3 show that the relationship is between Charlotte Lund Rasmussen and Andreas Holtermann, as both are from Denmark. The third largest group (in dark blue) is also composed of six authors, among whom Antonella Buccianti, Matar Thiombane, Benedetto de Vivo, and Stefano Albanese stand out. Looking at Table 3, one can see that all four of these authors are in Italy, and two of them—Matar Thiombane and Benedetto de Vivo—are affiliated with the University of Pegaso. The next cluster has four authors (in yellow), followed by Cluster 5 (dark violet), in which three authors are related to each other but do not have an established relationship with the previous groups, as well as Cluster 6 (light blue), which has three authors.

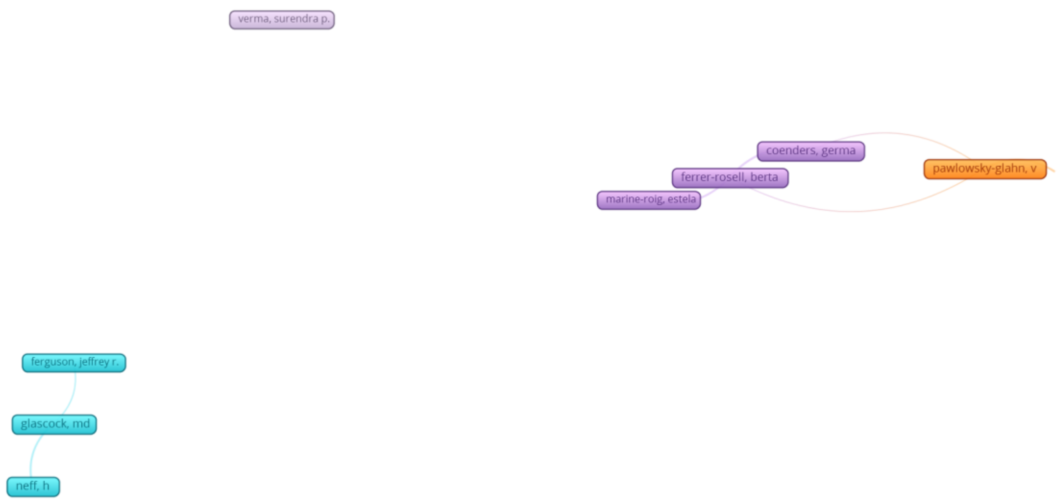

Figure 3. Coauthorship with a threshold of five documents and the 100 most representative connections
Source: VOSviewer software.

Independently, and without establishing relationships with the rest of the clusters, we observed that Cluster 7 has two authors, Juan Jose Egozcue and Vera Pawlowsky-Glahn. We should note that both authors have recognised prestige in the application of the CoDa methodology. Finally, there are the last eight clusters, with less importance in terms of the analysis of coauthorship.

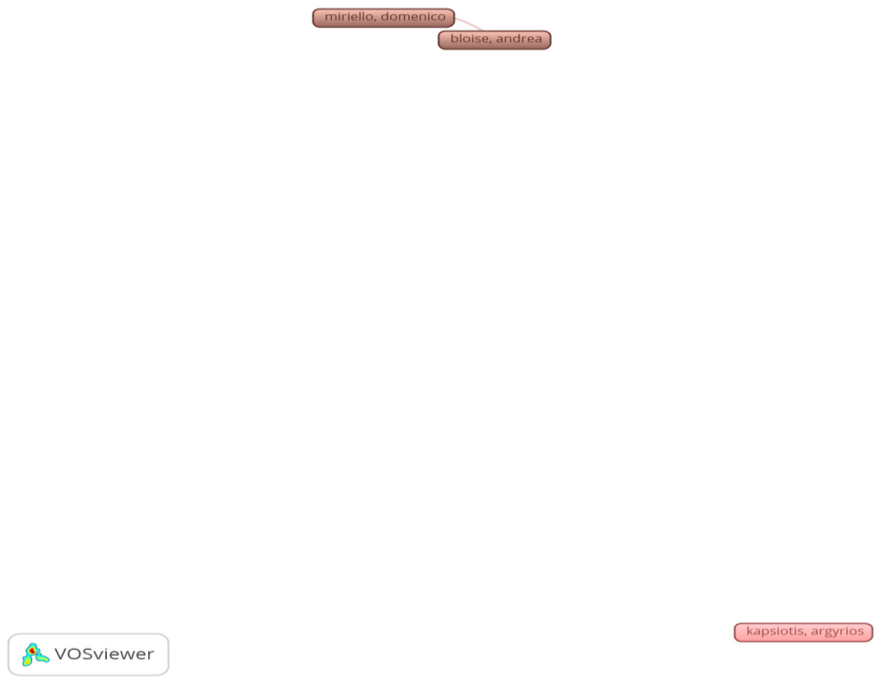

Figure 4. Coauthorship with a threshold of five papers and the 100 most representative connections
Source: VOSviewer software.

Finally, with some exceptions, one can see that most of the authors depicted in Figure 2 are also represented in Table 3, so there are no significant differences between the analysis performed using the WoS (total count) and the VOSviewer software (fractional count).

4.3. More Productive and Influential Institutions

In this section, we identify the 46 most productive institutions and the bibliographic linkage between them, together with these institutions' position according to the Academic Ranking of World Universities (ARWU; Shanghai Ranking Consultancy 2019) and the Quacquarelli Symonds World University Ranking (QS 2020), thus supplying an answer to RQ3.

Table 4. Most Productive Institutions in the Publication of Articles with CoDa Methodology in the Area of Social Sciences According to the Web of Science.

R	INSTITUTIONS	PAÍS	≥100	≥50	≥25	≥10	≥5	≥1	TP	TC	H	SC	C/P	ARWU	QS
1	University of Girona	Spain	0	5	13	24	32	40	42	814	16	61	19,38	801-900	-
2	The French National Centre for Scientific Research (CNRS)	France	0	2	6	17	24	31	32	536	13	2	16,75	-	-
3	University of South Australia	Australia	2	6	9	17	20	28	29	946	14	53	32,62	401-500	-
4	The National Research Council (Cnr)	Italy	2	3	7	19	22	27	28	855	15	10	30,54	-	-
5	UDICE French Research Universities	France	0	2	5	13	20	27	28	375	11	3	13,39	-	-
6	University of London	UK	2	6	11	15	17	24	27	976	14	5	36,15	-	-
7	Chinese Academy of Sciences	China	1	5	10	16	21	26	27	737	15	1	27,30	-	-
8	Helmholtz Association	Germany	0	5	8	16	20	23	23	552	12	3	24,00	-	-
9	China University of Geosciences	China	0	3	7	16	19	21	21	467	13	5	22,24	301-400	-
10	University of Missouri Columbia	Usa	0	1	4	10	13	18	21	270	10	2	12,86	201-300	476
11	University College London	UK	0	3	6	9	11	18	19	381	9	5	20,05	17	-
12	University of Cambridge	UK	0	0	4	10	12	16	18	257	10	6	14,28	3	3
13	University of Southern Denmark (SDU)	Denmark	0	0	2	6	12	16	18	188	8	16	10,44	301-400	309
14	University of Florence	Italy	0	0	5	11	15	17	17	307	10	10	18,06	201-300	451
15	United States Department of Energy (DOE)	Usa	1	2	2	5	12	17	17	305	7	0	17,94	-	-
16	University of Naples Federico II	Italy	0	0	2	13	14	17	17	256	11	19	15,06	301-400	424
17	Polytechnic University of Catalonia	Spain	0	1	3	3	8	14	17	189	7	17	11,12	701-800	319
18	James Hutton Institute	UK	1	4	6	10	14	16	16	769	10	20	48,06	-	-
19	Palacký University Olomouc	Czech republic	1	3	4	9	11	14	16	407	9	24	25,44	701-800	601-650
20	Ghent University	Belgium	0	0	4	8	10	15	16	223	9	10	13,94	71	141
21	Australian National University	Australia	0	0	2	6	11	16	16	194	8	1	12,13	76	27
22	Glasgow Caledonian University	UK	2	2	7	11	12	15	15	877	11	27	58,47	151-200	1001-1200
23	University of Ottawa	Canada	2	6	11	13	13	15	15	806	12	17	53,73	201-300	230
24	Russian Academy of Sciences	Russia	0	0	0	7	8	13	15	100	8	0	6,67	-	-
25	University of Alberta	Canada	1	2	9	9	11	13	14	512	9	5	36,57	101-150	126
26	National Autonomous University of Mexico	México	0	0	1	3	7	11	13	663	9	14	51,00	201-300	105
27	The University of Sydney	Australia	0	0	5	9	10	12	13	265	9	3	20,38	69	38
28	U.S Departament of Agriculture (USDA)	Usa	0	0	2	8	11	12	13	195	9	1	15,00	-	-
29	National Research Centre for the Working Environment (NFA)	Denmark	0	0	1	3	7	11	13	128	6	16	9,85	-	-
30	Children's Hospital of Eastern Ontario (CHEO)	Canada	2	5	10	11	11	12	12	739	11	16	61,58	-	-
31	University of Adelaide	Australia	2	4	6	11	11	12	12	511	10	5	42,58	101-150	108
32	University of Oxford	UK	1	3	7	9	11	11	12	451	9	7	37,58	7	2
33	The University of Melbourne	Australia	0	2	5	7	8	12	12	314	7	2	26,17	33	37
34	Universiy of Gavle	Sweeden	0	0	1	2	7	11	12	104	6	17	8,67	-	-
35	University of Barcelona	Spain	0	0	0	2	9	12	12	79	6	3	6,58	151-200	168
36	University of Washington	Usa	3	4	5	8	9	11	11	552	9	1	50,18	19	85
37	Imperial College London	UK	1	3	4	8	10	10	11	340	8	0	30,91	25	7
38	Arizona State University	Usa	1	1	2	7	10	10	11	331	8	0	30,09	101-150	216
39	University of Paris	France	0	1	2	6	9	11	11	188	7	1	17,09	73	261
40	National Aeronautics Space Administration (NASA)	Usa	0	1	2	5	9	11	11	169	7	0	15,36	-	-
41	California Institute of Technology (Caltech)	Usa	0	1	2	4	9	11	11	153	6	2	13,91	9	6
42	University of Helsinki	Finland	1	4	6	7	9	10	10	437	8	3	43,70	82	104
43	Commonwealth Scientific Industrial Research Organisation	Australia	1	2	2	4	8	9	10	365	6	0	36,50	-	-
44	University of Calabria	Italy	1	1	3	6	6	9	10	246	6	8	24,60	701-800	101-1200
45	University of North Carolina	Usa	0	0	4	6	8	10	10	192	8	0	19,20	29	95
46	University of Copenhagen	Denmark	0	0	0	2	6	8	10	76	5	9	7,60	30	76

Notes: ≥100, ≥50, ≥25, ≥10, ≥5, ≥1 = number of papers with at least 100, 50, 25, 10, 5, and 1 citations. CoDa = compositional data analyses; R = ranking; TP = total papers; TC = total citations; S/C = self-citations; H = h-index; C/P = citations per paper; ARWU 2019 = Academic Ranking of World
Source: Web of Science, June 2022.

The two most productive institutions are the University of Girona in Spain and The French National Centre for Scientific Research in France. In fact, three of the 51 most productive authors are located at the former. Only 10 institutions—two from China, two from France, and the rest from different countries—produced more than 20 papers.

When we analysed the number of citations, the ranking changed completely. In the first position we find the University of London (United Kingdom), followed by the University of South Australia, the Glasgow Caledonian University (United Kingdom), the National Research Council (Italy), and the University of Girona (Spain). The University of Ottawa (Canada), the James Hutton Institute (United Kingdom), and the Children's Hospital of Eastern Ontario (Canada), with a lower number of publications, are in Positions 4, 6, 7, and 8 if we order the ranking in Table 4 by a number of citations.

In regard to the h-index, we can see that all the institutions in Table 4 obtained an h-index ≥5, although 18 of the 46 institutions obtained an h-index ≥10. As for the institutions with papers that have received 100 or more citations, we can see that the University of Washington leads the ranking, with three times as many publications as its followers. In this respect, 19 institutions have a paper with 100 or more citations. The values increase if we analyse the institutions affiliated with papers with 50 or more citations because 31 institutions meet this criterion. It is worth noting that 19 institutions have obtained citations in all their publications, and two of them, Helmholtz Association and the China University of Geosciences, are in the top 10 in terms of total publications. On the other hand, 16 institutions have obtained citations in almost all of their publications (except one), and the remaining 10 institutions have between two and three uncited papers.

With regard to the countries represented in Table 4, we can see that the 46 institutions are distributed among 16 countries, with the United States being the country with the largest number of institutions, with a total of eight. This is followed by the United Kingdom, with seven institutions and, in third place, Australia, with six.

In terms of the ARWU ranking, 14 universities in Table 4 appear in the top 100, with the University of Cambridge occupying the highest position (Position 3). In total, 31 universities are represented in the ARWU ranking. Similarly, 10 universities appear in the top 100 of the QS ranking, with the University of Oxford occupying the highest position on this list (Position 2), followed again by the University of Cambridge (Position 3) and the California Institute of Technology (Position 9). In total, 27 universities are represented in the QS ranking.

We next graphically analysed the degree to which authors from the main institutions cited the same documents. We did this using the bibliographic linking technique. Figure 5 shows the results of this analysis, which had a minimum threshold of five documents and the top 100 connections.

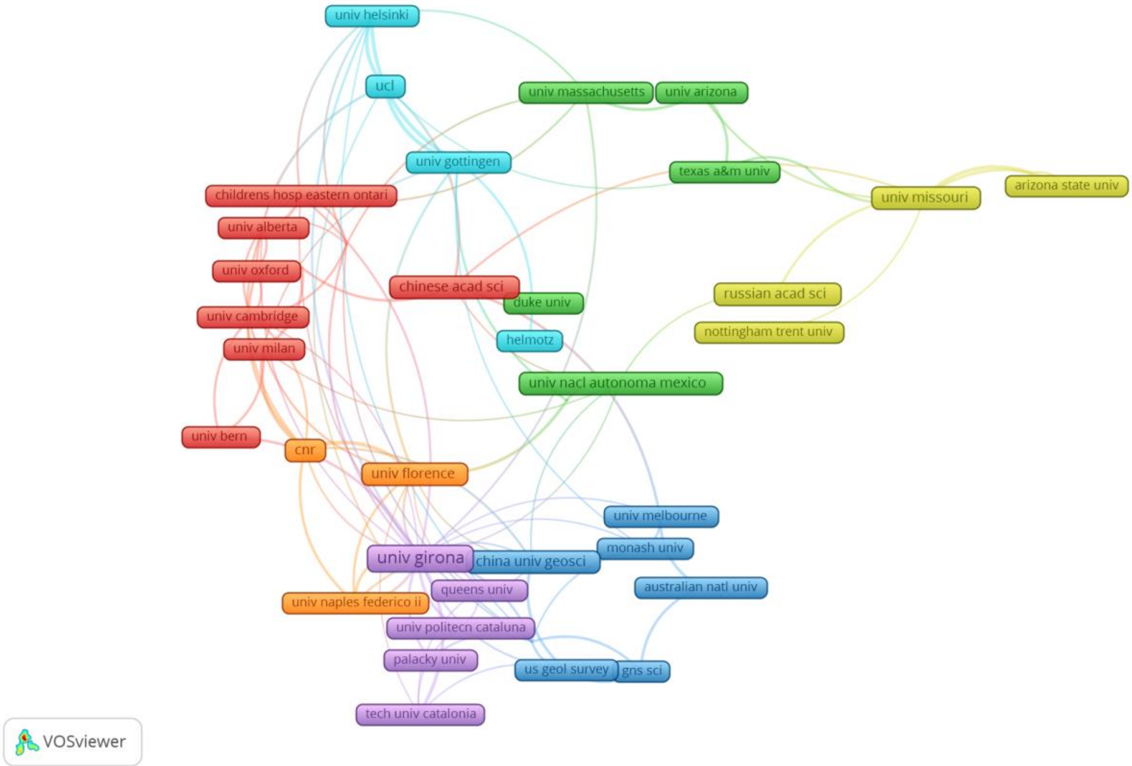

Figure 5. Bibliographic linkage with a threshold of five documents and the 100 most representative connections
Source: VOSviewer software.

Figure 5 shows five main nodes and three secondary nodes. The main node, in red, is made up of 10 institutions, including the University of Cambridge and the University of Oxford. It can also be seen that three of these institutions are in the United Kingdom. The second most important node, in green, is made up of eight institutions, many of them from the United States, such as Washington, Massachusetts, Arizona, and Texas. The third most important node, in blue, is made up of seven institutions, most of them from the Pacific area. The fourth most important node, in yellow, is made up of six institutions, most of them from Anglo-Saxon countries. The two secondary nodes are led by the University of Girona, the University of Florence, and the University of Helsinki.

Figure 5 shows how collaborations between institutions tend to occur in closer geographical areas, although the two leading institutions, the University of Surrey and Bournemouth University, also maintain important connections with the rest of the nodes. Finally, with some exceptions, one can see that most of the organisations in Figure 5 are also shown in Table 4, so there is no significant difference between the analyses performed on the basis of counting, addition, and fraction counting.

4.4. Most Productive and Influential countries

In this section, we identify the 35 most productive countries and the bibliographic linkage between them, supplying an answer to RQ4.

Table 5. Most Productive Countries in the Area of Social Sciences According to the Web of Science.

R	PAIS	≥ 100	≥ 50	≥ 25	≥ 10	≥ 5	≥ 1	TP	TC	H-INDEX	SC	Pop	TP/Pop	TC/Pop
1	USA	20	42	87	170	220	288	310	8.623	45	45	329,5	0,94	26,17
2	UK	9	33	77	125	149	185	196	5.921	43	210	67,22	2,92	88,08
3	Australia	3	14	28	63	82	116	121	2.540	26	101	25,69	4,71	98,87
4	China	3	14	34	58	80	103	118	2.708	30	22	1402	0,08	1,93
5	Italy	6	14	33	72	84	108	115	2.978	28	57	59,55	1,93	50,01
6	Spain	1	10	26	37	73	97	104	1.891	26	129	47,35	2,20	39,94
7	Germany	2	14	26	52	70	85	88	2.076	25	12	83,24	1,06	24,94
8	Canada	4	13	27	39	56	79	83	1.944	25	30	38,01	2,18	51,14
9	France	2	8	14	28	37	51	54	1.367	19	7	67,39	0,80	20,28
10	Belgium	1	2	13	22	26	34	35	714	16	11	11,56	3,03	61,76
11	Sweden	1	8	14	19	27	33	34	981	16	21	10,35	3,29	94,78
12	Netherlands	2	4	8	14	25	29	34	706	12	6	17,44	1,95	40,48
13	Brazil	0	2	6	15	16	27	31	404	12	3	212,6	0,15	1,90
14	Denmark	1	3	7	13	22	28	30	557	12	22	5,831	5,14	95,52
15	India	1	4	6	14	19	26	26	748	12	3	1408	0,02	0,53
16	Switzerland	1	3	7	15	18	24	24	556	12	2	8,64	2,78	64,35
17	Iran	0	1	6	9	17	24	24	344	9	2	83,99	0,29	4,10
18	Finland	2	5	9	11	16	20	21	726	11	12	5,53	3,80	131,28
19	Mexico	1	5	5	10	12	19	21	702	10	16	128,9	0,16	5,45
20	Greece	0	4	7	14	16	20	21	498	12	1	10,72	1,96	46,46
21	Czech Republic	1	3	5	11	13	18	21	458	11	24	10,7	1,96	42,80
22	New Zealand	1	4	9	12	14	19	20	574	12	2	5,08	3,94	112,99
23	Portugal	1	3	5	10	12	17	20	476	10	5	10,31	1,94	46,17
24	Japan	1	3	5	10	14	18	20	425	10	2	125,8	0,16	3,38
25	Russia	0	0	1	9	11	16	20	155	9	0	144,1	0,14	1,08
26	Norway	0	1	4	8	14	17	19	398	9	0	5,38	3,53	73,98
27	Austria	0	0	1	7	11	15	17	172	8	1	8,917	1,91	19,29
28	South korea	0	1	1	4	8	13	15	173	6	0	51,78	0,29	3,34
29	Ireland	1	5	9	12	13	13	14	643	11	1	4,99	2,81	128,86
30	South Africa	1	3	6	8	9	12	12	477	8	0	59,31	0,20	8,04
31	Turkey	0	2	2	6	10	12	12	201	7	1	84,34	0,14	2,38
32	Argentina	0	0	1	2	3	8	11	60	4	0	45,38	0,24	1,32
33	Chile	0	0	1	4	6	9	10	124	6	1	19,12	0,52	6,49
34	Croatia	0	1	1	4	6	10	10	124	6	2	4,047	2,47	30,64
35	POLAND	0	0	0	5	6	10	10	91	6	1	37,95	0,26	2,40

Notes: ≥100, ≥50, ≥25, ≥10, ≥5, ≥1 = number of papers with at least 100, 50, 25, 10, 5, and 1 citation; R ranking; TP = total papers; TC = total citations; S/C = self-citations; H = h-index; C/P = citations per paper; Pop = population in millions; TP/Pop = papers per inhabitant, in millions; TC/Pop = citations per inhabitant, in millions.
Source: Web of Science, June 2022.

Table 5 shows that the United States, the United Kingdom, and Spain are the most productive and influential countries in the publication of documents using the CoDa methodology in the area of Social Sciences. In the ranking, only the first seven countries have more than 100 documents. If we analyse the number of citations, the ranking related to the top 10 countries does not change significantly from the ranking related to the number of papers. The only difference we noticed is that Sweden is no longer in the top 10; its place has been taken by Belgium.

A similar situation occurs concerning the h-index. In this respect, 24 countries in Table 5 have two-digit h-indexes, with the United States in the first position, with a total of 45. As for the countries with papers with 100 or more citations, we can see how the United States again leads the ranking, obtaining more than twice as many publications as its immediate followers. In this respect, 23 countries have a paper with 100 or more citations. The values increase significantly if we analyse the countries with papers with 50 or more citations; in this case, 30 countries meet this criterion.

Finally, the countries India, Switzerland, South Africa, Iran, Turkey, Croatia, and Poland managed to cite all of their works. If we analyse productivity per million inhabitants, we can see that the ranking does change substantially, as it is now led by Denmark, Australia, and New Zealand in terms of a number of publications. In terms of the number of citations per million inhabitants, the ranking changes again, with Finland, Ireland, and New Zealand taking the top positions. It is noteworthy that the United Kingdom, Belgium, and Australia continue to hold a privileged position among the countries that receive the highest number of citations per million inhabitants, despite the fact that the United Kingdom was the 12th country in Table 4 with the highest number of inhabitants.

Finally, we graphically visualised the extent to which authors from the main countries cited the same documents. We did this using the bibliographic linkage technique. The results of this analysis, which had a minimum threshold of five documents and the top 100 connections, are shown in Figure 6.

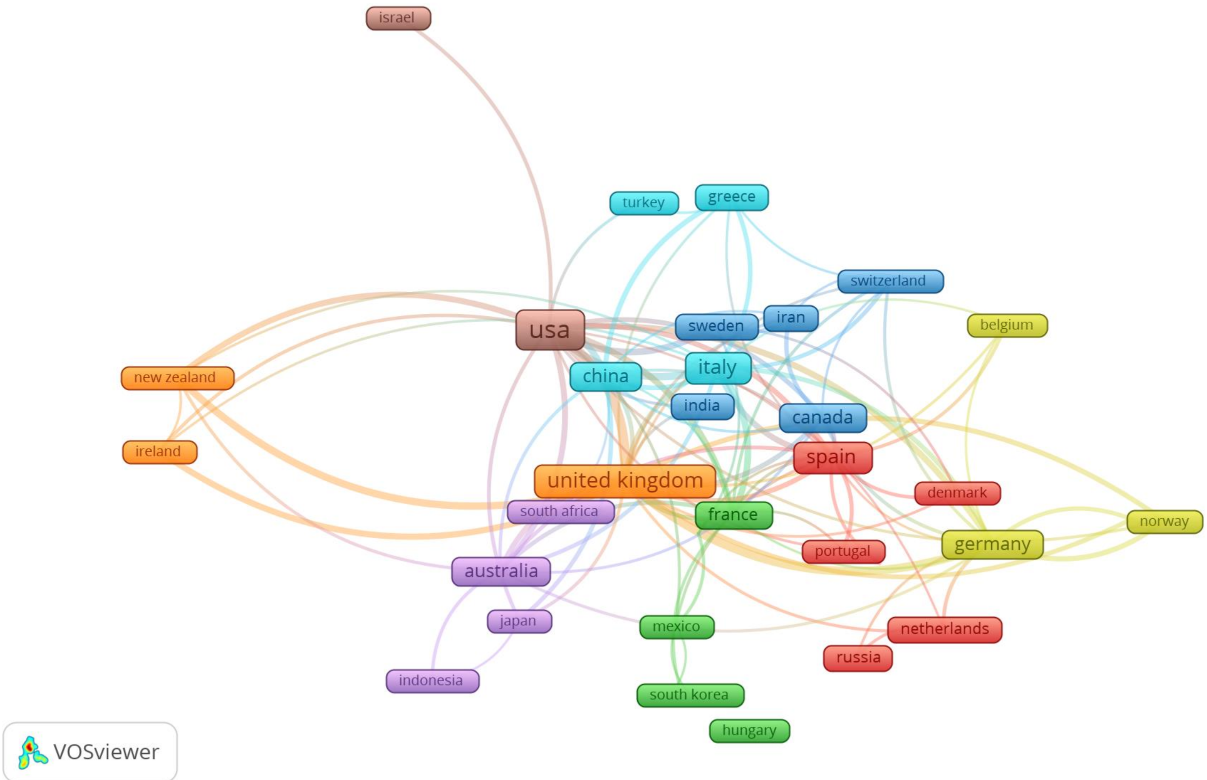

Figure 6. Bibliographic coupling with a threshold of five documents and the 100 most representative connections
Source: VOSviewer software.

Figure 6 depicts eight clusters. The red, green, and dark blue nodes, with six countries each, are formed by the countries with the highest number of connections in the area of the social sciences. Next are three nodes with four countries each, led by Belgium, China, and Australia, respectively. The last two nodes encompass three and two countries, respectively. Figure 6 clearly shows how there is a multitude of collaborations between countries, which are not limited to countries that are exclusively part of the same node. It is worth noting that the countries with the highest number of publications are found in the last two nodes.

We observed that the United Kingdom maintains relations with a multitude of countries in other clusters, such as China, India, Canada, South Africa, Germany, Norway, Australia, and New Zealand, among others. At the same time, we can see how the United States maintains relations and connections with a large number of countries, including the United Kingdom, France, Australia, Italy, and China.

As for global evolution over time, the following graph shows the evolution from 2012 to 2016.

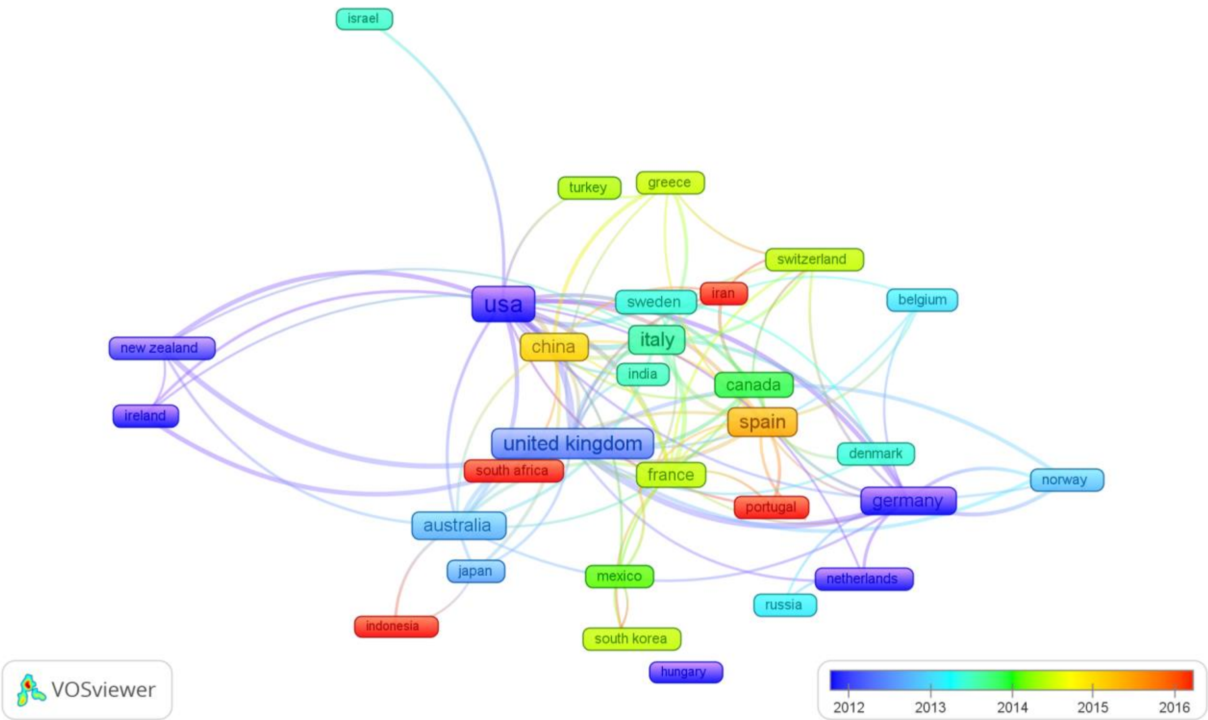

Figure 7. Temporal evolution of bibliographic coupling with a threshold of five documents and the 100 most representative connections
Source: VOSviewer software.

As shown in Figure 7, the countries with the longest tradition and the greatest weight in regard to the total number of publications began to publish from the beginning of this period, such as the United States, New Zealand, Germany, Hungary, and Ireland. As the colour changes to warmer shades, one can see how there are countries that have recently begun to have more productive activity, and these are the same countries that have relations and connections with the pioneers in this field. These countries have emerging authors in publishing research that has used the CoDa methodology in the social sciences, so it is to be expected that their output will grow. These countries include Indonesia, South Africa, Iran, and Portugal.

As with the previous graphic analyses, one can also see that most of the countries in Figure 6 are also represented in Table 5, with no significant differences between the analysis carried out using the WoS (total count) and the VOSviewer software (fractional count). Although we can indeed highlight the relationships established among the different countries, the analysis carried out through the VOSviewer shows that countries that are in the top positions by the number of publications, according to the total count, belong to clusters with fewer members.

5. Conclusions

Through a bibliometric analysis, we have analysed the main authors, institutions, and countries that apply the CoDa methodology in the area of social sciences. Because assessing the quality of published papers is neither easy nor obvious, we used a wide range of bibliometric methods and indicators.

To answer RQ1, we analysed the structure of publications and citations in the field of social sciences by authors who used the CoDa methodology. The citation structure revealed that 92% of the published papers have been cited at least once. Furthermore, the citation structure shows that the number of publications per year has increased significantly since 2007.

For RQ2, we analysed the 51 most productive authors and their coauthorship structure. The most productive author is Dorothea Dumuid. She also had the highest h-index and number of citations per article, indicating that many of her articles are highly cited. In terms of the highest number of citations, the first position goes to Sebastien F.M, Chastin. The second most productive author is Timothy Olds. Finally, the results also show that the number of authors affiliated with institutions in the Czech Republic is more relevant than in other countries.

To answer RQ3, we have analysed the 46 most productive institutions and the bibliographic coupling that occurs among them, together with their position according to the ARWU and QS rankings. Of the 46 universities, the two most productive institutions are the University of Girona in Spain and The French National Centre for Scientific Research in France. As for the ARWU ranking, 14 universities appear in the top 100 and, in total, 31 are represented in the ARWU ranking. Similarly, 10 universities appear in the top 100 of the QS ranking and, in total, 27 are represented in the QS ranking.

Finally, for RQ 4, we analysed the 35 most productive countries and the bibliographic links among them. The United States is in first place, followed by the United Kingdom and Australia. Bibliographic cohesion shows that cultures and regions tend to cooperate between countries that are close to each other either by geographical or cultural proximity.

This study has some limitations. The first is related to the database used to carry out this work, the WoS. In this sense, most of the journals included in this analysis have been integrated into the WoS since 2007, so the analysis of the first years included only a few journals. The second limitation is that we did not consider all documents indexed in the WoS; only those that had undergone a rigorous refereeing process were analysed. Finally, the WoS collects data with a full-count system, a limitation that we tried to overcome by using partial counting with the VOSviewer software. The analysis explained bibliographic unions and coauthors under fractional counts. In this sense, the results obtained with the full count and the partial count were very similar, so all the results obtained in our work are reliable.

While acknowledging these limitations, we believe that this work is sufficiently rigorous to provide an overview of the research context of CoDa methodology in the social sciences. In addition, this study has some practical implications, including that the results obtained will help social science scholars identify the authors, institutions, and countries most likely to develop and share research results (Law and Chon, 2007). Finally, as Mulet-Forteza et al. (2019) showed, this study allows researchers to compare the present results with those of other studies in the field of social sciences, enabling them able to judge new trends through experimental methods that facilitate the development of new theories that can lead to the formulation of new conceptual frameworks.

6. References

Aguillo IF (2012) Is Google Scholar useful for bibliometrics? A webometric analysis. *Scientometrics 91*(2):343-351. https://doi.org/10.1007/s11192-011-0582-8

Aitchison J (1982) The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological) 44*(2):139-160.

Alonso-Gamboa JO, Reyna-Espinosa FR (2015) Revistas académicas mexicanas: Panorama y prospectiva. *CIENCIA ergo-sum: revista científica multidisciplinaria de la Universidad Autónoma del Estado de México 22*(3):181-191.

Baas J, Schotten M, Plume A, Côté G, Karimi R (2020) Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. Quantitative Science Studies 1(1):377-386. https://doi.org/10.1162/ qss_a_00019

Batista-Foguet JM, Ferrer-Rosell B, Serlavós R, Coenders G, Boyatzis RE (2015) An alternative approach to analyze ipsative data. Revisiting experiential learning theory. *Frontiers in Psychology 6*:1742. https://doi.org/10.3389/fpsyg.2015.01742

Benckendorff, P. (2009). Themes and trends in Australian and New Zealand tourism research: A social network analysis of citations in two leading journals (1994-2007). *Journal of Hospitality and Tourism Management*, *16*(1), 1-15.

Beigel F (2013a) David y Goliath. El sistema académico mundial y las perspectivas del conocimiento producido en la periferia. *Revista Pensamiento Universitario 15*:1-18.
Beigel F (2013b) The politics of Academic Autonomy in Latin America, Ashgate, London.

Beigel F (2014) Publishing from the periphery: Structural heterogeneity and segmented circuits. The evaluation of scientific publications for tenure in Argentina's CONICET. *Current Sociology 62*(5):743-765. https://doi.org/10.1177/0011392114533977

Beigel F (2016) Científicos Periféricos, entre Ariel y Calibán. Saberes Institucionales y Circuitos de Consagración en Argentina: Las Publicaciones de los Investigadores del CONICET. *Revista Dados 59*(4):215-255. http://dx.doi.org/10.1590/001152582017136

Benckendorff P, Zehrer A (2013) A network analysis of tourism research. *Annals of Tourism Research 43*:121-149. https://doi.org/10.1016/j.annals.2013.04.005

Belles-Sampera J, Guillen M, Santolino M (2016) Compositional methods applied to capital allocation problems. *The Journal of Risk 19*(2):15-30.

Birkle C, Pendlebury DA, Schnell J, Adams J (2020) Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies 1*(1):363-376. https://doi.org/10.1162/qss_a_00018

Blasco-Duatis M, Sáez Zafra M, García Fernández N (2018) Compositional representation (CoDa) of the agenda-setting of the political opinion makers in the main Spanish media groups in the 2015 general election. *Communication & Society 31*(2):1-24. https://doi.org/10.15581/003.31.2.1-23

Bornmann L (2014) How are excellent (highly cited) papers defined in bibliometrics? A quantitative analysis of the literature. *Research Evaluation 23*:166-173. https://doi.org/10.1093/reseval/rvu002

Bornmann L, Leydesdorff L (2014) Scientometrics in a changing research landscape: bibliometrics has become an integral part of research quality evaluation and has been changing the practice of research. *EMBO reports 15*(12):1228-1232. https://doi.org/10.15252/embr.201439608

Brueton VC, Vale CL, Choodari-Oskooei B, Jinks R, Tierney JF (2014) Measuring the impact of methodological research: A framework and methods to identify evidence of impact. Trials 15:464. https://doi.org/10.1186/1745-6215-15-464

Carreras Simó M, Coenders G (2020) Principal component analysis of financial statements. A compositional approach. Revista de Metodos Cuantitativos para la Economía y la Empresa 29:18-37. https://doi.org/10.46661/revmetodoscuanteconempresa.3580

Carreras-Simó M, Coenders G (2021) The relationship between asset and capital structure: A compositional approach with panel vector autoregressive models. Quantitative Finance and Economics 5(4):571-590. https://doi.org/10.3934/QFE.2021025

Castanha R, Gracio M (2014) Bibliometrics Contribution to the Metatheoretical and Domain Analysis Studies. Knowledge Organization 41(2):171-174. https://doi.org/10.5771/0943-7444-2014-2-171

Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F (2011) An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Informetrics 5*(1):146-166. https://doi.org/10.1016/j.joi.2010.10.002

Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F (2011) Science mapping software tools: review, analysis and cooperative study among tools. Journal of the American Society for Information *Science and Technology 62*:1382-1402. https://doi.org/10.1002/asi.21525

Coenders G, Ferrer-Rosell B (2020) Compositional data analysis in tourism. Review and future directions. *Tourism Analysis 25*(1):153-168. https://doi.org/10.3727/108354220X15758301241594

Corona Berkin S (2017) Flujos metodológicos desde el Sur latinoamericano. La zona de la comunicación y las Metodologías Horizontales. *Comunicación y Sociedad 30*:69-106. https://doi.org/10.32870/cys.v0i30.6819

Chakraborty S, Majumdar AS, Shukla AD (2020) Role of fluid in strain softening within the main central thrust in Sikkim: The origin of quartz-rich mylonites. *Journal of Structural Geology 140*:104-145. https://doi.org/10.1016/j.jsg.2020.104145

Daunis-i-Estadella J, Thió-Henestrosa S, Mateu-Figueras G (2011) Including supplementary elements in a compositional biplot. *Computational Geosciences 37*(5):696-701. https://doi.org/10.1016/j.cageo.2010.11.003

Ding Y, Rousseau R, Wolfram D (2014) Measuring scholarly impact: Methods and practice. Springer.

Diodato V (1994) Dictionary of bibliometric. Haworth Press.

Ezbakhe F, Pérez Foguet A (2020) Child mortality levels and trends: A new compositional approach. *Demographic Research 43*:1263-1296. https://dx.doi.org/10.4054/DemRes.2020.43.43

Falagas ME, Pitsouni EI, Malietzis GA, Pappas G (2008) Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *The FASEB Journal 22*(2):338-342. https://doi.org/10.1096/fj.07-9492lsf

Ferrer-Rosell B, Coenders G, Martínez-Garcia E (2015) Determinants in tourist expenditure composition—The role of airline types. *Tourism Economics 21*(1):9-32. https://doi.org/10.5367/te.2014.0434

Filzmoser P, Hron K, Templ M (2018) Applied compositional data analysis with worked examples in R. Springer.

Gehanno JF, Rollin L, Darmoni S (2013) Is the coverage of Google Scholar enough to be used alone for systematic reviews. BMCMedical *Informatics and Decision Making 13*(1):7. https://doi.org/10.1186%2F1472-6947-13-7

Giustini D, Boulos MN (2013) Google Scholar is not enough to be used alone for systematic reviews. *Online journal of public health informatics 5*(2): 214. https://doi.org/10.5210%2Fojphi.v5i2.4623

Gläser J, Laudel GA (2015) Bibliometric reconstruction of research trails for qualitative investigations of scientific innovations. *Historical Social Research 40*:299-330. http://dx.doi.org/10.12759/hsr.40.2015.3.299-330

Godin B (2006) On the origins of bibliometrics. *Scientometrics 68*(1):109-133. https://doi.org/10.1007/s11192-006-0086-0

Graffelman J, Pawlowsky-Glahn,V, Egozcue JJ, Buccianti A (2018) Exploration of geochemical data with compositional canonical biplots. *Journal of Geochemical Exploration 194*:120-133. https://doi.org/10.1016/j.gexplo.2018.07.014

Greenacre M (2018) Compositional data analysis in practice. Chapman and Hall/CRC press.

Gumpenberger C, Wieland M, Gorraiz J (2012) Bibliometric practices and activities at the University of Vienna. *Library Management 33*:174-183. https://doi.org/10.1108/01435121211217199

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. https://doi.org/10.1007/s11192-018-2958-5

Haddaway NR, Collins AM, Coughlin D, Kirk S (2015) The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PLOS One 10*(9):e0138237. https://doi.org/10.1371/journal.pone.0138237

Halevi G, Moed H, Bar-Ilan J (2017) Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation - Review of the Literature. *Journal of Informetrics 11*(3):823- 834. https://doi.org/10.1016/j.joi.2017.06.005

Harzing AW (2016) Microsoft Academic (Search): A Phoenix arisen from the ashes? *Scientometrics 108*(3):1637-1647. http://dx.doi.org/10.1007/s11192-016-2026-y

Harzing AW, Alakangas S (2016) Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics 106*(2):787-804. https://doi.org/10.1007/s11192-015-1798-9

Hron K, Coenders G, Filzmoser P, Palarea-Albaladejo J, Faměra M, Matys Grygar T (2021) Analysing pairwise logratios revisited. *Mathematical Geosciences 53*:1643-1666. https://doi.org/10.1007/s11004-021-09938-w

Jacso P (2005) As we may search – comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citationenhanced databases. *Current Science 89*(9):1537-1547.

Jiang J, Fan L (2022) Visualizing the knowledge domain of language experience: A bibliometric analysis. *Sage Open 12*(1):1-17. https://doi.org/10.1177/21582440211067252

Kessler MM (1963) Bibliographic Coupling between Scientific Papers. *American Documentation 14*(1):10-25. https://doi.org/10.3390/environments9020028

Kogovšek T, Coenders G, Hlebec V (2013) Predictors and outcomes of social network compositions. A compositional structural equation modeling approach. *Social Networks 35*(1):1-10. https://doi.org/10.1016/j.socnet.2012.10.002

Köseoglu MA, Mehraliyev F, Aladag OF, King B (2021) Origins, evolution and themes of scholarly hospitality sources: 1960-2019. *International Journal of Hospitality Management 94*:102817. https://doi.org/10.1016/j.ijhm.2020.102817

Koseoglu MA, Mehraliyev F, Xiao H (2019) Intellectual connections in tourism studies. *Annals of Tourism Research 79*:1-15. https://doi.org/10.1016/j.annals.2019.102760

Kousha K, Thelwall M (2008) Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines. *Scientometrics 74*(2):273-294. http://dx.doi.org/10.1007/s11192-008-0217-x

Law R, Chon K (2007) Evaluating research performance in tourism and hospitality: The perspective of university program heads. *Tourism Management 28*(5):1203-1211. https://doi.org/10.1016/j.tourman.2006.09.022

Linares-Mustarós S, Coenders G, Vives-Mestres M (2018) Financial performance and distress profiles. From classification according to financial ratios to compositional classification. *Advances in Accounting 40*:1-10. https://doi.org/10.1016/j.adiac.2017.10.003

Lu Y, Wang Y, Li B, Li J, Jiang HQ (2020) Temporal and spatial variations in haze research: a bibliometric analysis. *Environmental Review 1*(28):12-20. https://doi.org/10.1139/er-2018-0112

López-Cózar ED, Orduna-Malea E, Martín-Martín A (2019) Google Scholar as a data source for research assessment. In Glaenzel W, Moed H, Schmoch U, Thelwall M (Eds.), Springer handbook of science and technology indicators, Springer, Berlin.

Ma, R. (2012). Author bibliographic coupling analysis: A test based on a Chinese academic database. *Journal of Informetrics*, *6*(4), 532-542.

Martín-Martín A, Orduna-Malea E, Ayllón JM, Delgado López-Cozar E (2016) A two-sided academic landscape: snapshot of highly-cited documents in Google Scholar (1950-2013). *Revista española de documentación científica 39*(4):1-21. https://doi.org/10.3989/redc.2016.4.1405

Martín-Martín A, Orduna-Malea E, Thelwall M, López-Cózar E (2018) Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics 12*(4):1160- 1177. https://doi.org/10.1016/j.joi.2018.09.002

McBurney MK, Novak PL (2002) What is bibliometric and why should you care? [Conference session]. Proceedings of the Professional Communication Conference, Portland, OR (pp. 108-114).

Martorell Cunill O, Socias Salvá A, Otero Gonzalez L, Mulet-Forteza C (2019) Thirty-fifth anniversary of the International Journal of Hospitality Management: A bibliometric overview. *International Journal of Hospitality Management 78*:89-101. https://doi.org/10.1016/j.ijhm.2018.10.013

Martyn J (1964) Bibliographic coupling. *Journal of Documentation 20*(4):236. https://doi.org/10.1108/eb026352

Mauleon-Mendez E, Genovart-Balaguer J, Merigo J, Mulet-Forteza C (2018) Sustainable Tourism Research Towards Twenty-Five Years of the Journal of Sustainable Tourism. *Advances in Hospitality and Tourism Research (AHTR) 6*(1):23-46. http://dx.doi.org/10.30519/ahtr.406555

Mauleón-Méndez E, Genovart-Balaguer J, Martorell-Cunill O, Mulet-Forteza C (2020) Tourism research: A bibliometric and country analysis. *Journal of Intelligent & Fuzzy Systems 38*(5):5565-5577. https://doi.org/10.3233/JIFS-179647

Meho LI, Yang K (2007) Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology 58*(13):2105-2125. https://doi.org/10.1002/asi.20677

Merigó JM, Cancino CA, Coronado F, Urbano D (2016) Academic research in innovation: A country analysis. *Scientometrics 108*(2):559-593. https://doi.org/10.1007/s11192-016-1984-4

Merigó JM, Gil-Lafuente AM, Yager RR (2015). An overview of fuzzy research with bibliometric indicators. *Applied Soft Computing 27*:420-433. https://doi.org/10.1016/j.asoc.2014.10.035

Merigó JM, Mulet-Forteza C, Valencia C, Lew AA (2019) Twenty years of tourism geographies: A bibliometric overview. *Tourism Geographies 21*(5):881-910. https://doi.org/10.1080/14616688.2019.1666913

Merigó JM, Blanco-Mesa F, Gil-Lafuente AM, Yager R (2017) Thirty Years of the International Journal of Intelligent Systems: A Bibliometric Review. *International Journal of Intelligent Systems 32*(5):526-554. https://doi.org/10.1002/int.21859

Mikki S (2010) Comparing Google Scholar and ISI Web of Science for earth sciences. *Scientometrics 82*(2):321-331. https://doi.org/10.1007/s11192-009-0038-6

Mikšová D, Filzmoser P, Middleton M (2020) Imputation of values above an upper detection limit in compositional data. *Computational Geosciences 136*:104383. https://doi.org/10.1016/j.cageo.2019.104383

Moed HF, Bar-Ilan J, Halevi G (2016) A new methodology for comparing Google Scholar and Scopus. *Journal of Informetrics 10*(2):533-551. https://doi.org/10.1016/j.joi.2016.04.017

Mosbah-Natanson S, Gingras Y (2015) The globalization of social sciences? Evidence from a quantitative analysis of 30 years of production, collaboration and citations in the social sciences (1980-2009). *Current Sociology 62*(5):626-646. https://doi.org/10.1177/0011392113498866

Mulet-Forteza C, Genovart-Balaguer J, Mauleon-Mendez E, Merigó JM (2019) A bibliometric research in the tourism, leisure and hospitality fields. *Journal of Business Research 101*:819-827. https://doi.org/10.1016/j.jbusres.2018.12.002

Mulet-Forteza C, Socias Salvá A, Monserrat S, Amores A (2020) 80th anniversary of pure and applied geophysics: A bibliometric overview. *Pure and Applied Geophysics 177*:531-570. https://doi.org/10.1007/s00024-019-02328-8.

Mulet-Forteza, C., Lunn, E., Merigó, J.M., Horrach Rosselló, P. (2021). Research Progress in Tourism, Leisure & Hospitality in Europe (1969-2018). *International Journal of Contemporary Hospitality Management*, *33*(1), 48-74.

Muller I, Hron K, Fiserova E, Smahaj J, Cakirpaloglu P, Vancakova J (2018) Interpretation of compositional regression with application to time budget analysis. *Austrian Journal of Statistics 47*(2):3-19. https://doi.org/10.17713/ajs.v47i2.652

Navarro-Lopez C, Gonzalez-Morcillo S, Mulet-Forteza C, Linares-Mustaros S (2021) A bibliometric analysis of the 35th anniversary of the paper" the statistical analysis of compositional data" by John Aitchison (1982). *Austrian Journal of Statistics 50*(2):38-55. http://dx.doi.org/10.17713/ajs.v50i2.1066

Neuhaus C, Neuhaus E, Asher A, Wrede C (2006) The depth and breadth of Google Scholar: An empirical study. Portal: *Libraries and the Academy 6*(2):127-141. http://dx.doi.org/10.1353/pla.2006.0026

Noyons E, Moed H, Van Raan A (1999) Integrating research performance analysis and science mapping. *Scientometrics 46*(3):591-604. https://doi.org/10.1007/BF02459614

Orduña-Malea E, Martín-Martín A, Ayllón JM, López-Cózar ED (2016) La revolución Google Scholar: Destapando la caja de Pandora académica. UNE, Granada.

Ortells R, Egozcue JJ, Ortego MI, Garola A (2016) Relationship between popularity of key words in the Google browser and the evolution of worldwide financial indices. In Martín-Fernández JA, Thió-Henestrosa S (Eds.), Compositional data analysis. Springer proceedings in mathematics & statistics (p. 187). Springer.

Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2015) Modeling and analysis of compositional data (p. 2015). Wiley.

Perianes-Rodriguez A, Waltman L, van Eck NJ (2016) Constructing bibliometric networks: A comparison between full and fractional counting. *Journal of Informetrics 10*(4):1178-1195. https://doi.org/10.1016/j.joi.2016.10.006

Pospiech S, Tolosana-Delgado R, van Den Boogaart KG (2021) Discriminant analysis for compositional data incorporating cell-wise uncertainties. *Mathematical Geosciences 53*:1-20. https://doi.org/10.1007/s11004-020-09878-x

Pritchard A (1969) Statistical bibliography or bibliometrics? *Journal of Documentation 25*:348-349.

Rodrigues LA, Daunís-l-Estadella J, Mateu-Flgueras G, Thió-Henestrosaê S (2011) Flying in compositional morphospaces: Evolution of limb proportions in flying vertebrates. In Pawlowsky-Glahn V, Buccianti A (Eds.) Compositional data analysis: Theory and applications (pp. 235-254) John Wiley & Sons, Ltd.

Sanz-Sanz E, Martinetti D, Napoléone C (2018) Operational modelling of peri-urban farmland for public action in Mediterranean context. *Land Use Policy 75*:757-771. https://doi.org/10.1016/j.landusepol.2018.04.003

Shafique M (2013) Thinking inside the box? Intellectual structure of the knowledge base of innovation research (1988-2008). Strategic *Management Journal 34*(1):62-93. https://doi.org/10.1002/smj.2002

Svensson G (2010) SSCI and its impact factors: A 'prisoner's dilemma'? *European Journal of Marketing 44*(1-2):23-33. https://doi.org/10.1108/03090561011008583

Thomas CW, Aitchison J (2005) Compositional data analysis of geological variability and process: A case study. *Mathematical Geology 37*(7):753-772. https://doi.org/10.1007/s11004-005-7378-4

Tolosana-Delgado R, Mueller U, van Den Boogaart KG (2019) Geostatistics for compositional data: An overview. *Mathematical Geosciences 51*(4):485-526. https://doi.org/10.1007/s11004-018-9769-3

Torraco RJ (2016) Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review 15*(4):404-428. https://doi.org/10.1177/1534484316671606

Van Den Boogaart KG, Tolosana-Delgado R (2013) Analyzing compositional data with R. Springer.

Van Eck NJ, Waltman L (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics 84*(2):523-538. https://doi.org/10.1007/s11192-009-0146-3

Verma SP, Guevara M, Agrawal S (2006) Discriminating four tectonic settings: Five new geochemical diagrams for basic and ultrabasic volcanic rocks based on log—Ratio transformation of major-element data. *Journal of Earth System Science 115*(5):485-528. https://doi.org/10.1007/BF02702907

Vogel R (2014) What happened to the public organization? A bibliometric analysis of public administration and organization studies. *The American Review of Public Administration 44*:383-408. https://doi.org/10.1177/0275074012470867

Zhao D, Strotmann A (2008) Evolution of research activities and intellectual influences in information science 1996-2005: Introducing author bibliographic-coupling analysis. *Journal of the American Society for Information Science and Technology 59*(13): 2070-2086. https://doi.org/10.1002/asi.20910

Zupic I, Čater T (2015) Bibliometric methods in management and organization. *Organizational Research Methods* *18*(3):429-472. https://doi.org/10.1177/1094428114562629

10. Final Conclusions

Based on the review of the different publications in the different categories represented in the Web of Science from 1982 to 2022, a method of bibliometric analysis was proposed, through evaluative techniques and relational techniques, to help researchers to know the evolution in the application of the compositional data analysis (CoDa) methodology. In this way, the method presented in this doctoral thesis aims to help researchers develop their future bibliometric research through the application of CoDa methodology, providing information on new research areas, possible future collaborations and hot topics.

The conclusions obtained in the three articles are associated with bibliometric research on the application of CoDa methodology and based on the documents published by John Aitchinson (1982, 1986), as well as its specific application in the social sciences field, and will be stated below.

In the first article, **A bibliometric analysis of the 35th anniversary of the paper "The Statistical Analysis of Compositional Data" by John Aitchison (1982)**, an exhaustive bibliometric analysis of the aforementioned paper was carried out, which allowed us to respond to all the objectives established in our first work. Specifically, we answered the four research questions posed at the beginning. Concerning the first question, 'What is the evolution of the number of citations of Aitchison's 1982 paper?', we analysed how the number of citations of the paper evolved. The results showed that the paper has been cited continuously since its publication and that, in the last four years, the number of citations has increased more significantly, in line with the typical exponential growth expected. It has also been corroborated that almost 95% of the 784 citations received by Aitchison's (1982) article have come from papers that have undergone a strict refereeing process. As for the second research question, 'Who are the authors that most cite Aitchison's paper?', Vera Pawlowsky-Glahn and Glòria Mateu-Figueras (University of Girona), Antonella Buccianti (Università degli Studi di Firenze), Juan José Egozcue (Polytechnic University of Catalonia) and Raimon Tolosana-Delgado (HZDR – Helmholtz-Zentrum Dresden-Rossendorf) are the authors who most cited Aitchison's 1982 paper. In addition, authors from the University of Girona, the Polytechnic University of Catalonia, the University of Florence, the Helmholtz Association and the Center National de La Recherche Scientifique (CNRS) are the authors who have most cited the paper, thus answering the third research question, 'What are the institutions that most cite Aitchison's paper?', while the fourth research question, 'What are the countries that most cite Aitchison's paper?' showed that authors from the United States, United Kingdom, Spain, Italy, Australia and Germany are those who have cited the paper the most.

Regarding the second article, **The statistical analysis of compositional data by John Aitchison (1986): A bibliometric overview**, bibliometric analysis was carried out in all the publications that have cited the book entitled 'The Statistical Analysis of Compositional Data' published by John Aitchison in 1986. As in the previous work, all the objectives set out therein were answered. As for the first research question, 'What is the academic structure of Aitchison's 1986 book?', the analysis provided information related to the citation structure of this work, showing that there has been a very significant increase in citations over the last four years. The temporal analysis also revealed that, in recent years, the CoDa methodology has mostly been used practically in scientific works, although it is also true that during the period 2011–2019 the theoretical formulation of this methodology started to be discussed. This is due to the significant increase in the number of publications in fields related to the 'Earth Sciences'. Regarding the second objective, 'In which main journals have Aitchison's book been cited?', we concluded that most journals were indexed in the WoS categories 'Geosciences, Multidisciplinary' and 'Ecology', specifically highlighting the journals 'Mathematical Geosciences' and 'Computers Geosciences'. Regarding the third research question, 'What main topics are analysed in the principal papers published by authors citing Aitchison's book?', our results show that the keywords of the main papers citing the book corresponded to the year 2010, coinciding with the period in which Aitchison's book obtained the highest number of citations. The results show that the main keywords were related to the field of earth sciences, among which we can highlight 'Biodiversity', 'Geodiversity', 'Geological Heritage' and 'Geological Resources'. By way of summary, our work has presented several findings that allow us to understand the evolution and advances that are occurring in the field of CoDa through the analysis of citations of Aitchison's book (1986), a seminal text in the field of geoscience as well as modern science.

The third article, **Research progress in compositional data in social sciences. A bibliometric analysis**, analyses the main authors, institutions and countries that apply the CoDa methodology in the social sciences, using a wide range of bibliometric methods and indicators. This responds to the different objectives and research questions posed. To answer the first research question, 'What is the academic structure of the application of CoDa methodology in the field of social sciences?', the structure of publications and citations in the field of social sciences by authors who used CoDa methodology was analysed. Identifying the structure of citations, and the origin and evolution of citations help identify intellectual links in academic fields applying CoDa in social sciences (Köseoglu et al., 2019; Shafique, 2013). Moreover, the work covers a long period, which allows researchers to get a complete picture of the area covered as well as its possible development. As for the second research question, 'What are the most productive authors in applying CoDa methodology in the area of social sciences?', we identified the 51 most productive authors, as well as their co-authorships. The most productive author was Dorothea Dumuid. Furthermore, this author obtained the highest H-index and the highest number of citations per article, which also shows that many of her articles were highly cited. As for the third research question, 'What are the most productive institutions in publishing articles with CoDa methodology in the area of social sciences?', we determined the 46 most productive institutions and the bibliographic coupling that occurred between them. Finally, for the fourth research question, 'Which are the countries that have published the most papers with compositional data analysis methodology applied in the area of social sciences?', we determined the 35 most productive countries and the bibliographic links between them. The United States was in first place, followed by the United Kingdom and Australia. In this respect, mapping intellectual connections contributes to the creation of new theories and the development of existing ones, providing insight into future directions that scientific research may take (Köseoglu et al., 2021). Therefore, the academic analysis of these developments will help researchers to determine the potential effects that the theory might have on society. In addition, these processes also provide valuable information for researchers and experts (Torraco, 2016), with a study that helps deepen their understanding of the current state of CoDa research, especially for those researchers who are not very familiar with this methodology (Jiang & Fan, 2022).

The achievement of all the objectives set out in the different works that make up this doctoral thesis allows us to affirm that the objectives set out in the thesis have been met.

11. Main results and discussion of this finding

Finally, in the last stage, the results obtained from the bibliometric analyses carried out must be described, interpreted and discussed, with the main purpose of advancing the field of application of compositional data analysis methodology, as well as proposing new research lines.

When we talk about bibliometrics we must see that it is a discipline that, using mathematical and statistical methods, provides data on different aspects of scientific literature. Therefore, we complement the evaluation of the quality of scientific publications through their analysis and, as is well known, this takes on special importance in the university world where quality is an increasingly important factor when it comes to considering the different processes such as accreditation and recognition of research areas, among others. In any area where studies and research are generated, it is important to measure them, as this is the way to identify the key fundamentals of a specific topic and the possibilities for future research lines.

Compositional data analysis is a methodology, it could be said, of recent application and where it is necessary to evaluate the fields in which it is being applied to know the evolution that the methodology has undergone in the different scientific disciplines in which it has been applied. Thus, the use of bibliometric techniques will be useful to evaluate future research in different institutions, as well as a whole set of related entities, such as governors and decision-makers at a macro level with more complete results, as well as to journal evaluators when it comes to understanding this methodology and having a global understanding of the effectiveness, efficiency, production and quality of the publications that use it.

Therefore, this doctoral thesis has carried out an analysis of three aspects that are essential to provide a realistic body of the evaluation and use of this technique in the different scientific categories indexed in WoS. The first step was to analyse the publications that had cited John Aitchison's paper (1982) to understand the evolution of this publication. Subsequently, a bibliometric analysis was made of the book published by John Aitchison (1986), The Statistical Analysis of Compositional Data, where a more detailed explanation was given of how the methodological application could be carried out. In both cases, exponential growth can be seen in all scientific categories. As for the third article, a specific evaluation was carried out in the area of social sciences. Therefore, the third publication aims to review how the production, efficiency and research quality of those publications that were carried out with the compositional data analysis methodology in the field of social sciences has increased.

It can be seen how the recent application of the methodology based on compositional data analysis has provided the opportunity for studies to address the needs and objectives most demanded and needed by today's society and by institutions at the international level, such as the 17 Sustainable Development Goals (SDGs) of 2030 of the United Nations.

Having analysed the top 20 publications in the 12 most important WoS categories in terms of the total number of publications applying CoDa, it can be seen how research is carried out at the micro, meso and macro levels of sociology. Thus, from a micro and meso level, we find studies such as, for example, the association of arsenic metabolism with cancer, vascular diseases and diabetes, epidemiology in different parts of the planet, and the study of chemical and mineralogical characterisation providing information on the origin and composition of elements used by ancient tribes which facilitates their organisational knowledge, characteristics of specific populations, regional studies of air pollution, indications for small and medium-sized enterprises for the application of Big Data, the dynamics of litter on Mediterranean coastal beaches, the development of new elements that can function as biodiesel (rubber seeds), methods for monitoring and evaluating the prevention of household waste, the empowerment of women as a facilitating factor in the use of contraceptives in sub-Saharan Africa, how to understand the spending patterns and volume of low-cost airline users, among many other examples. As can be seen, these studies are linked to the sustainable development goals related to health, health and well-being, gender equality, affordable and clean energy, sustainable cities and communities, and responsible production and consumption.

At the macro level, issues such as sustainability and air pollution, greenhouse gas reduction during the life cycle of electric vehicles, plastic waste recycling by analysing the most polluting elements, and the effects of the global pandemic on people who have had to work from home and the effects on 24-hour time use in office workers, nutritional assessment of food, biodiversity and nutrition, a common path towards global food security and sustainable development, an early warning system for banking failures based on novel localised pattern learning and semantically associative fuzzy neural network, adiposity and isotemporal substitution of physical activity, sedentary time and sleep among school children, human development index and children’s health-related quality of life and movement behaviours, human rights through NGOs and foreign aid delivery, among some examples. As can be seen, these studies are linked to the sustainable development goals related to health, health and well-being, decent work and economic growth, climate action, a life of terrestrial ecosystems, affordable and clean energy, sustainable cities and communities, responsible production and consumption.

On the other hand, as a relevant contribution, we highlight the publication (Appendix) where the connection between the application of the compositional data analysis (CoDa) methodology and bibliometric analysis is established through relational techniques. This publication was presented at the 9th International Workshop on Compositional Data Analysis (CoDaWork2022) held in Toulouse.

This doctoral thesis has several limitations. The first relates to the database used to carry out the studies, i.e. WoS, since, as we have indicated above, this database only includes a small number of academic journals, although all of them are considered relevant and influential. Another limitation of this database is that it uses a 'complete count' method to collect data. To address this limitation, our research also incorporated the 'fractional count' method, using VOSviewer software. The third limitation is that only those papers published in WoS that have passed a strict refereeing process have been considered. While acknowledging these limitations, we believe that this PhD thesis is sufficiently rigorous to provide an overview of the research context of compositional data analysis (CoDa) methodology. In addition, the findings of this dissertation are useful for authors to have a quick snapshot of what is happening, and what is expected to happen, in the research fields applying compositional data analysis methodology.

The doctoral thesis allows us to propose future lines of research. Firstly, an overview of the literature on the development of compositional data analysis (CoDa) research has been provided through a bibliometric analysis. This will enable future researchers to identify research gaps where this methodology is not yet applied (Faruk et al., 2021). This thesis also provides a starting point for future studies, as our results can be complemented with those obtained in other journals that choose to be included in the Emerging Sources Citation Index, as these journals offer less experienced researchers a good opportunity to publish their results. This may also lead to the development of emerging themes and new research trends (Mulet-Forteza et al., 2019) in the field of CoDa through the comparison of results, which, in turn, may lead to the development of new conceptual frameworks (Mulet-Forteza et al., 2021). It may also be interesting to repeat this work using different databases, such as Scopus, and not only WoS, to compare the results obtained. In addition, the methodology used in this study can be used to obtain similar results in other contexts than those proposed in this doctoral thesis (Quintero-Quintero et al., 2021).

Despite all of the above, it should be borne in mind that bibliometric studies are not a substitute for literature reviews. Bibliometrics can reliably connect publications, authors or journals, identify research sub-transmissions and produce maps of published research, but it is up to the researcher and his or her knowledge of the field to interpret the findings obtained.

12. References

Acedo, F.J., Barroso, C., Casanueva, C. & Galán, J.L. (2006) Co-Authorship in Management and Organizational Studies: An Empirical and Network Analysis. *Journal of Management Studies, 43* (5), 957-983

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological), 44*(2), 139-160.

Aitchison, J. (1986). *The statistical analysis of compositional data. Monographs on statistics and applied probability*. London: Chapman and Hall.

Barceló-Vidal, C., & Martín-Fernández, J.A. (2016). The mathematics of compositional analysis. *Austrian Journal of Statistics, 45*(4), 57-71.

Benckendorff P, Zehrer A (2013) A network analysis of tourism research. *Annals of Tourism Research 43*:121-149. https://doi.org/10.1016/j.annals.2013.04.005

Bornmann L, & Leydesdorff L. (2014). Scientometrics in a changing research landscape. *Science & Society, 15*,1228-1231.

Boyack, K.W., Klavans, R. & Börner, K. (2005) Mapping The Backbone of Science. *Scientometrics, 64*, 351-374

Boyack KW, Klavans R (2014). "Creation of a Highly Detailed, Dynamic, Global Model and Map of Science." Journal of the Association for Information Science and Technology, 65(4), 670–685.

Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics, 5*(1), 146-166.

Ding, K. & Tan, Y. (2015) Attract-repulse fireworks algorithm and its CUDA implementation using dynamic parallelism. *International Journal Swarm Research, 6* (2), 1-31.

Faruk, M., Rahman, M., & Hasan, S. (2021). How digital marketing evolved over time: A bibliometric analysis on scopus database. *Heliyon, 7*(12), e08603.

Gläser, J., & Laudel G. A. (2015). Bibliometric reconstruction of research trails for qualitative investigations of scientific innovations. *Historical Social Research*, *40*, 299-330.

Gumpenberger, C., Wieland, M., & Gorraiz, J. (2012). Bibliometric practices and activities at the University of Vienna. *Library Management*, *33*, 174-183.

Jiang, J., & Fan, L. (2022). Visualizing the Knowledge Domain of Language Experience: A Bibliometric Analysis. *Sage Open*, *12*(1). DOI:10.1177/21582440211067252.

Kessler MM (1963). “Bibliographic Coupling between Scientific Papers.” *American Documentation*, *14*(1), 10–25.

Köseoglu, M. A., Mehraliyev, F., & Xiao, H. (2019a). Intellectual connections in tourism studies. *Annals of Tourism research*, *79*, 1-15.

Köseoglu, M. A., Okumus, F., Dogan, I.C., & Law, R. (2019b). Intellectual structure of strategic management research in the hospitality management field: a co-citation analysis. *International journal of Hospitality* Management, *78*, 234-250.

Köseoglu, M. A., Mehraliyev, F., Aladak, O. F., & King, B. (2021). Origins, evolution and themes of scholarly hospitality sources: 1960–2019. *International journal of Hospitality Management*, *94*, 102817.

Leydesdorff, L. & Liwen, L. (2006) Co-occurrence matrices and their applications in informatic science: Extending ACA to the Web environment. *Journal of the American Society for Information Science and Technology, 57* (12), 1616-1628.

Liu, L., Jiang, T., Zhou, J., Mei, Y., Li, J., Tan. J., Wei, L., Li, J., Peng, Y., Chen, C., Liu, NN., Wang, H. (2022) Repurposing the FDA-approved anticanter agent ponatinib as a fluconazole potentiator by suppression of multidrug efflux and Pma1 expression in a broad spectrum of yeast species. *Microb Biotechnol*, *15*(2), 482-498

Lu, K. & Wolfram, D. (2012) Measuring author research relatedness: a comparison of word-based, topic-based and author cocitation approaches. *Journal of the American Society for Information Science and Technology 63* (10), 1973-1986.

McCain KW (1990). “Mapping Authors in Intellectual Space: A Technical Overview.” *Journal of the American Society for Information Science, 41*(6), 433.

Merigó, J. M., Gil-Lafuente, A. M., & Yager, R. R. (2015). An overview of fuzzy research with bibliometric indicators. *Applied Soft Computing*, *27*, 420-433.

Miguel, S.E., Caprile, L. Jorquera-Vidal, I. (2008) Análisis de co-términos y de redes sociales para la generación de mapas temáticos. *El Profesional de la Información, 17*(6), 637-646.

Moed, F., De Bruin, R.E., Nederhof, A.J., Van Raan, R.J.W., & Tijssen, A.F.J. (1992). *State of the art of bibliometric macro-indicators: An overview of demand and supply*. Science 149.

Mulet-Forteza C, Genovart-Balaguer J, Mauleon-Mendez E, Merig´o JM (2019a). “A Bib- liometric Research in the Tourism, Leisure and Hospitality Fields.” *Journal of Business Research, 101,* 819–827.

Mulet-Forteza C, Genovart-Balaguer J, Merig´o JM, Mauleon-Mendez E (2019b). “Bibliometric Structure of IJCHM in Its 30 Years.” *International Journal of Contemporary Hospitality Management, 31*(12), 4574–4604.

Nerur, S., Rasheed, A.A. & Pandey, A. (2015) Citation footprints on the sands of time: An analysis of idea migration in strategic management. *Strategic Management Journal, 37*(6), 1065-1084

Pearson K (1897). "Mathematical Contributions to the Theory of Evolution on a Form of Spurious Correlation which May Arise when Indices Are Used in the Measurement of Organs." *Proceedings of the Royal Society of London, 60*, 489–498.

Pilkington, A. & Lawton, T. (2014) Divided by a common language? Transnational insights into epistemological and methodological approaches to strategic management research in English-speaking countries. *International Journal of Strategic Management, 47* (5), 299-311

Quintero-Quintero, W., Blanco-Ariza, A.B., & Garzon-Castrillon, M.A. (2021). Intellectual Capital: A Review and Bibliometric Analysis. *Publications*, *9*(4), 46.

Racherla, P. & Hu, C. (2010) A social network perspective of tourism research collaborations. *Annals of Tourism Research, 37* (4), 1012-1034.

Ronda-Pupo, G.A.; Guerras-Martín, L.A. (2012) Dynamics of the evolution of the strategy concept 1962-2008: a co-word analysis. *Strategic Management Journal, 33*, 162-188.

Shafique, M. (2013). Thinking inside the box? Intellectual structure of the knowledge base of innovation research (1988–2008). *Strategic Management Journal, 34*(1), 62-93.
Schmidt, F. (2008). Meta-Analysis: A Constantly Evolving Research Integration Tool. *Organizational Research Methods*, *11*(1), 96-113.

Torraco, R. J. (2016). Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review, 15*(4), 404-428.

Vogel, R. (2014). What happened to the public organization? A bibliometric analysis of public administration and organization studies. *The American Review of Public Administration, 44,* 383-408.

White, H.D. & Griffith, B.C. (1981) Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science 32* (3), 163-171

White, H.D. & McCain (1998) Visualizing a discipline: and author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Sciemce, 49* (4), 327-355.

Ye, Q., Li, T. & Law, R. (2013) A coauthorship network analysis of tourism and hospitality research collaboration. *Journal of Hospitality & Tourism Research, 37* (1), 51-76.

Zupic I, Cater T (2015). Bibliometric Methods in Management and Organization. *Organizational Research Methods, 18*(3), 429–472.

APPENDIX

Since the publication of the seminal work on compositional data by John Aitchison in the *Journal of the Royal Statistical Society. Series B (Methodological)* in 1982, research in compositional data has expanded to a multitude of scientific disciplines. Through a bibliometric study, the aim of this paper is to update, as of May 2022, the data published in 2021 in the *Austrian Journal of Statistics* (Navarro-Lopez et al., 2021) regarding the impact that this seminal work has had on the scientific literature, updating the data concerning the main authors, institutions and research areas that have contributed to the expansion of knowledge in compositional data made by Aitchison in his 1982 publication. In relation to the analysis carried out with respect to the research areas, a biplot of the evolution of the fields of knowledge of the citing articles, for periods of 10 years, has also been carried out.

The bibliometric approach used includes both evaluative and relational techniques, with the aim of mapping the intellectual structure as well as the scientific progress that has been made since Aitchison's seminal work (1982). Among the evaluative techniques, indicators such as the number of publications and the number of citations have been used. The former is a measure of productivity, while the latter is a measure of influence (Svensson, 2010). These evaluative techniques have made it possible to determine the main authors, institutions and countries that have contributed to the expansion of the results of Aitchison's (1982) publication through different areas of research. As for the relational techniques, the VOSviewer software was used to map the relationships and connections between the main authors, institutions and countries that have cited Aitchison's (1982) work. Specifically, the relational techniques used were co-citation analysis and bibliographic linkage. Co-citation assumes that there is a relationship between two documents that have been cited jointly by a third document, while bibliographic coupling occurs when two documents include a common reference to a third document, so there is a possibility that these documents are linked. These two techniques make it possible to determine the academic structure of a scientific discipline.

First, the temporal evolution of the number of papers that have cited Aitchison's work (1982) has been analyzed. The results obtained show that the article has received citations uninterruptedly since its publication in 1982, although during its first years (1983-2007) the number of citations per year was very low, not exceeding 10 citations per year in any year (Figure 1). In contrast, since 2008, the number of annual citations has grown exponentially, reaching, during the year 2021, more than 350 citations. The biplot in Figure 2 shows the evolution of the fields of knowledge of the cited articles, by ten-year periods (Figure 2). The first ten years highlight the fields of clinical medicine, computer science and chemistry. The relative importance of mathematics and geosciences is greater in the period 1993-2002. The following period is dominated by agricultural sciences and physics, and the last by economics and business. In absolute terms and over the entire 40-year period, the fields with the most citations, in decreasing order, are clinical medicine, biology-biochemistry, mathematics, engineering and chemistry.

As for the authors who have contributed most to making Aitchison's (1982) work known in the scientific literature, Vera Pawlowsky-Glahn (University of Girona, Spain) is the researcher who has cited Aitchison's (1982) article the most, followed by Antonella Buccianti (Università degli Studi di Firenze, Italy) and Juan José Egozcue (Universidad Politécnica de Cataluña, Spain). Between these three authors, Aitchison's work (1982) has been cited a total of 108 times. It is remarkable to note that a total of 42 authors have cited Aitchison's (1982) publication 5 or more times. The graphical mapping, through an author-citation analysis, revealed four main types of collaborations (Figure 3). The first of these, consisting of seven authors, focuses on the figures of Dorothea Dumuid and Tim Olds, both from the University of South Australia. The second contains 5 authors focusing on the figures of Vera Pawlowsky-Glahn (University of Girona), Juan José Egozcue (Polytechnic University of Catalonia), Glòria Mateu-Figueras (University of Girona) and John Aitchison (University of Glasgow). This is the group with the largest network of connections, both between themselves and with the authors of the other nodes. The third group, consisting of four authors, is led by Javier Palarea-Albaladejo (Biomathematics & Statistics Scotland), while the fourth group, also with four authors, is led by Domenico Miriello and Andrea Bloise, both from the University of Calabria.

As for the main institutions that have cited the work of Aitchison (1982), the results reveal that the University of Girona is the one that leads the ranking, with more than 65 publications citing the work of Aitchison (1982), followed by the Centre Nationale de Reccherche Cientifique (CNRS) and the Helmhzolt Association, both with more than 50 papers that have cited this document. As for the graphical mapping, performed through a bibliographic linkage analysis, four groups of main nodes and five groups of secondary nodes have been revealed (Figure 4). The largest cluster, with 14 institutions, focuses on English-speaking institutions, including Harvard University, Duke University and the University of California at San Diego. The second cluster contains 13 institutions, including the University of South Australia, Victoria University and the University of Zurich. The third is composed of 11 institutions, including the University of Sao Paulo, the University of Turin and the University of Bremen, while the fourth main node, with nine institutions, is led by the University of Girona, the Polytechnic University of Catalonia, and the University of Florence. In turn, this last group of institutions is the one with the largest network of connections, both between themselves and with institutions in the other nodes, indicating that the authors of these institutions have led the expansion of the knowledge on compositional data realized by Aitchison in his 1982 publication.

This work is also not free of some limitations, the most important of which is related to the database used to carry out the study. Thus, since the data were collected from the Web of Science (WoS) database, the limitations of this database also apply to our analysis. The WoS database collects information under a "complete count" method, which means that papers with many coauthors generally carry more weight than papers produced by a single author (Mulet-Forteza et al., 2020). To address this limitation, we also used the "fractional count" method, using the VOSviewer software. No major differences were found to result from the choice of method, so that the results obtained, both from one method and the other, can be considered consistent and rigorous, so that the overview presented of Aitchison's 1982 article, which recently celebrated its 40th anniversary of publication, can be considered correct.

References

Aitchison, J. (1982). The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *44*(2), 139-160.

Mulet-Forteza, C., Socias Salvá, A., Monserrat, S., & Amores, A. (2020). 80th anniversary of pure and applied geophysics: A bibliometric overview. *Pure and Applied Geophysics*, *177*, 531–570.

Navarro, C., Gonzalez-Morcillo, S., Mulet-Forteza, C., & Linares-Mustaros, S. (2021). A bibliometric analysis of the 35th anniversary of the paper "the statistical analysis of compositional data" by John Aitchison (1982). *Austrian Journal of Statistics*, *50*(2), 38–55.

Svensson, G. (2010). SSCI and Its Impact Factors: A Prisoner's Dilemma? *European Journal of Marketing*, *44*, 23-33.

Figures

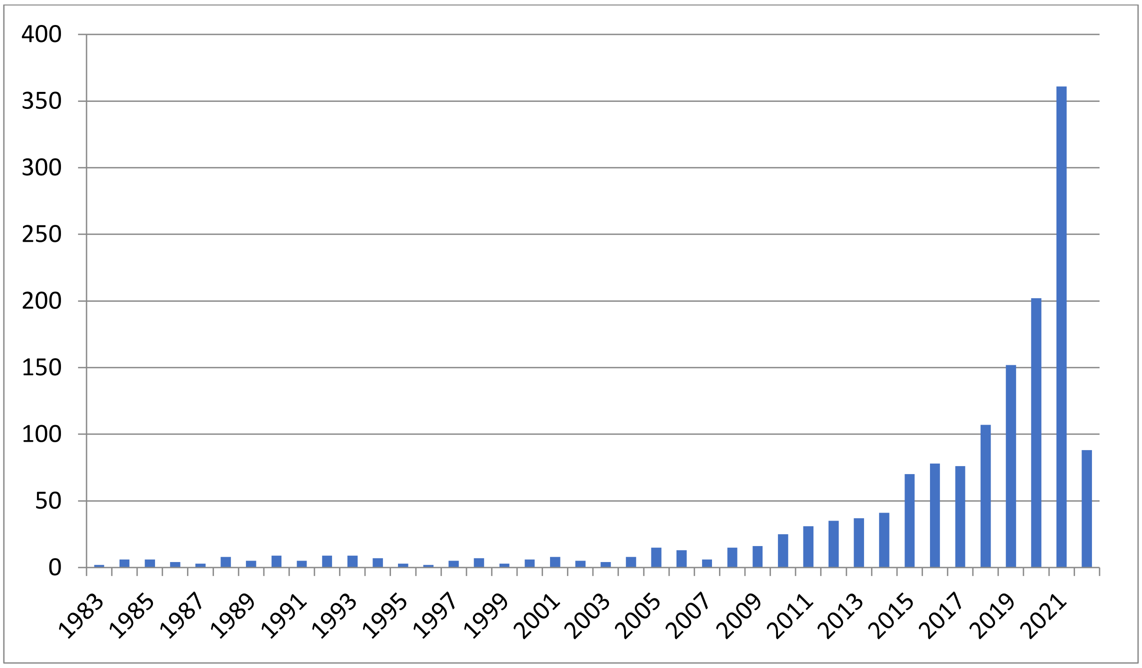

Figure 1: Annual number of citations received by Aitchison's 1982 paper. Source: own elaboration, compiled from WoS database.

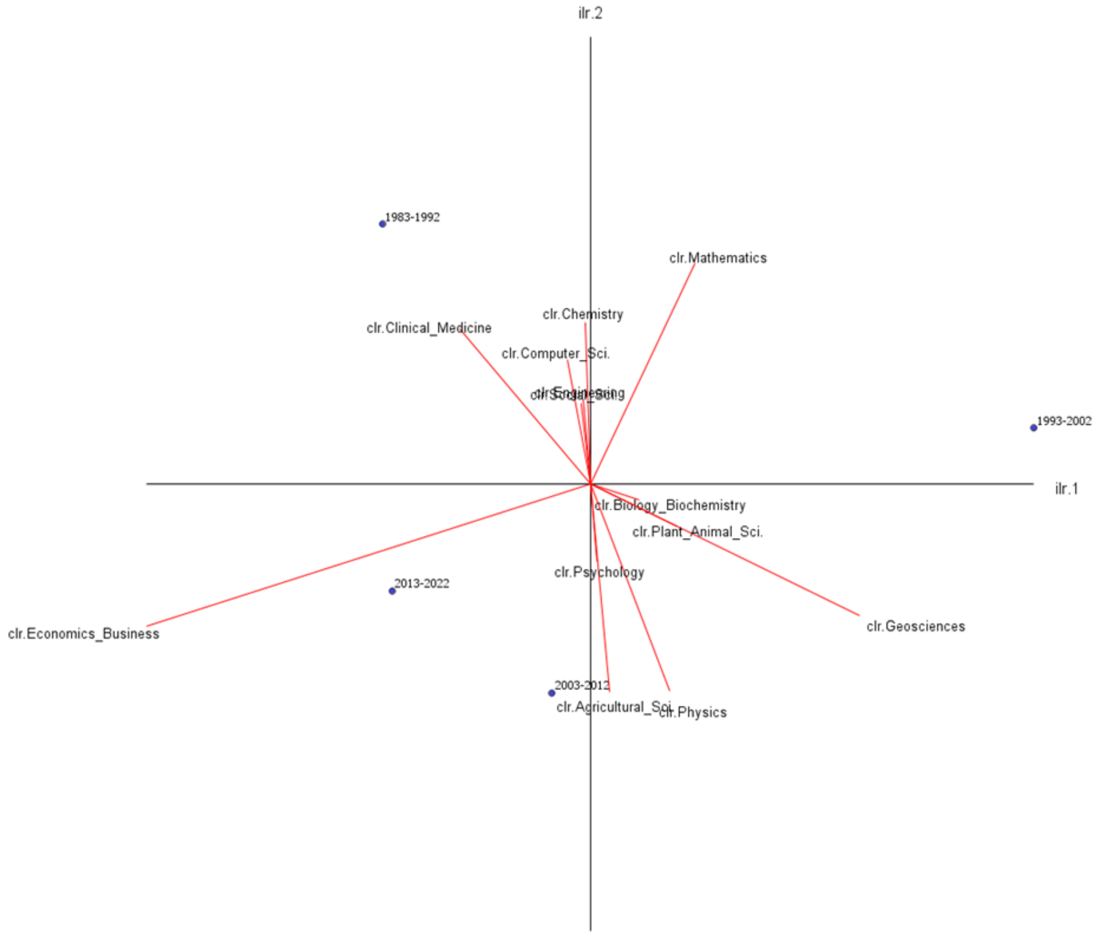

Figure 2: Application of the CoDa methodology in the different scientific categories considered in the Web of Science by decade.

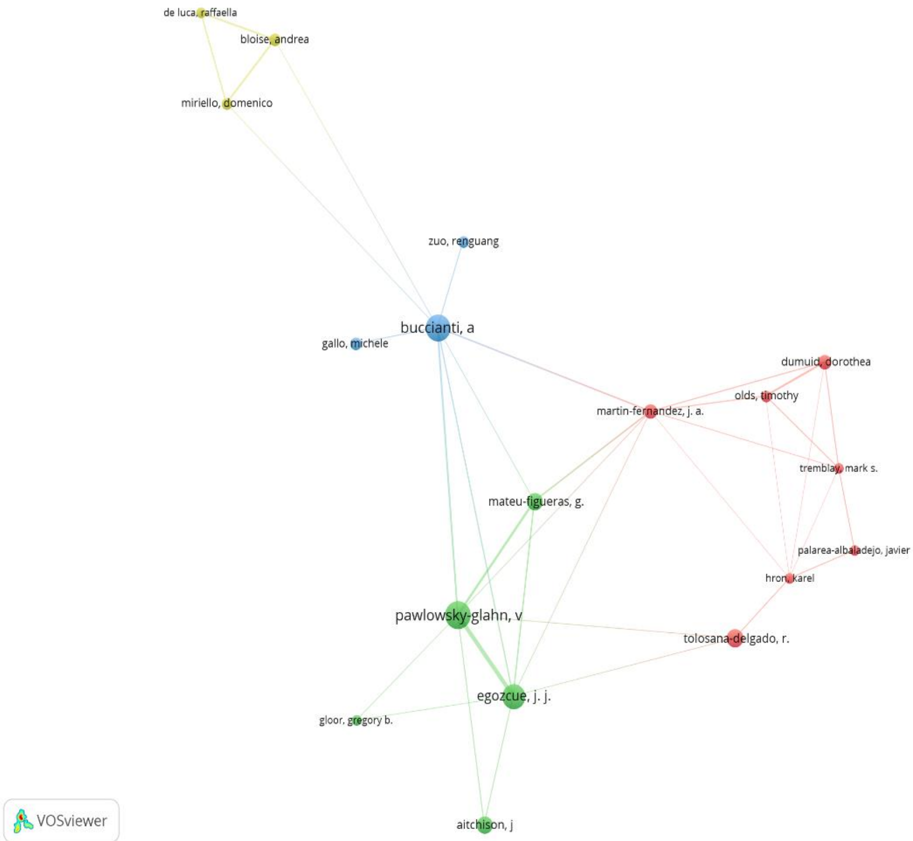

Figure 3: Co-citation of authors who have cited Aitchison's paper (1982). Node size = the number of citations received by an author; line thickness indicates multiple connections; line length is not significant; the colours highlight those authors which are a linked by a greater number of co-citations. Citation threshold of 5 and showing the 100 most representative co-citation connections. Source: own elaboration, based on WoS database; figure created using VOSviewer Software.

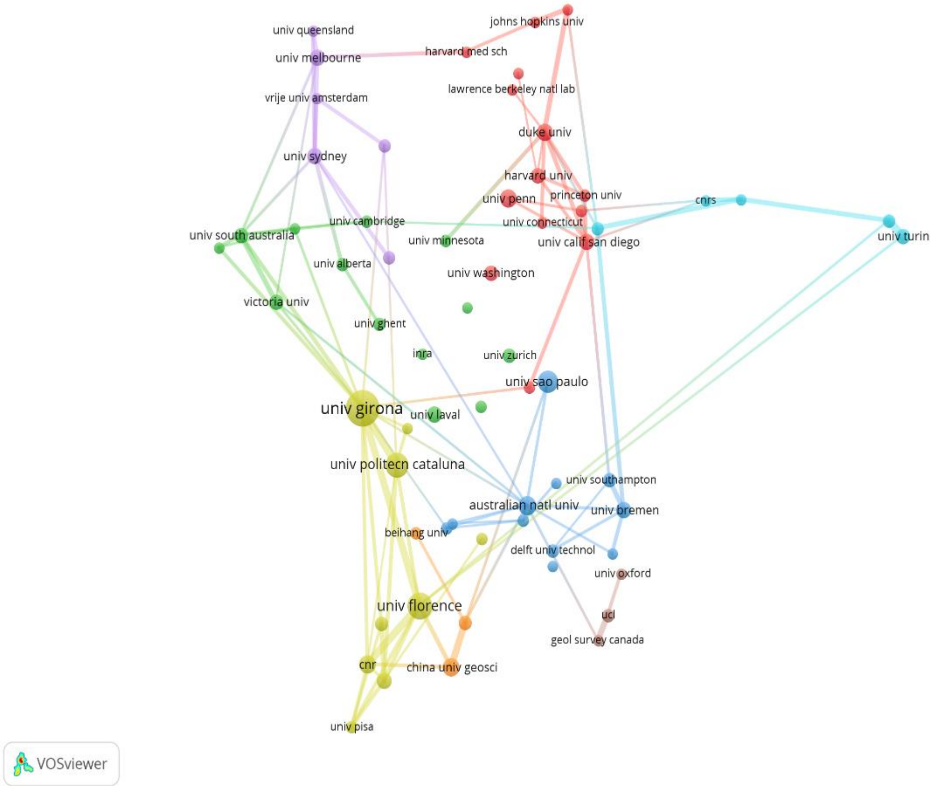

Figure 4: Bibliographic coupling of institutions citing Aitchison's paper (1982). Node size = number of citations received by authors belonging to a university; line thickness indicates multiple connections; line length is not significant; the colours highlight those institutions which are a linked by a greater number of co-citations. Citation threshold of five and showing the 100 most representative co-citation connections. Source: own elaboration, based on WoS database; figure created using VOSviewer Software.