Research Evaluation: Mapping the Field Structure

Avaliação da Pesquisa: Mapeando a Estrutura do Campo

Isabel Pinho[1] - [1]University of Aveiro | DCSPT | Aveiro | Aveiro | Portugal. Contato: isabelpinho@ua.pt. ORCID: https://orcid.org/ 0000-0003-1714-8979

Cláudia Pinho[2] - [2]University of Aveiro | DCSPT | Aveiro | Aveiro | Portugal. Contato: claudiapinho@ua.pt. ORCID: https://orcid.org/ 0000-0001-6249-5019

Maria J. Rosa[3] - [3]University of Aveiro | DEGEIT | Aveiro | Aveiro | Portugal. Contato: m.joao@ua.pt. ORCID: https://orcid.org/0000-0003-2956-6801

Resumo: O tema Avaliação da Pesquisa tem sido objeto de intenso estudo tanto pela sua importância para múltiplos propósitos, quanto pelo desenvolvimento teórico do modelo sistema de avaliação numa perspetiva de procura de melhoria contínua da qualidade de desempenho da pesquisa e do seu impacto societal e impacto na inovação. O objetivo deste estudo exploratório é recolher e mapear a estrutura intelectual da Avaliação de Pesquisa. Usando citações de redes, capturamos os blocos de construção latentes e as linhas condutoras deste tema. A visualização e análise de dados foram úteis para fazer uma exploração dos antecedentes da literatura; observando as redes de citação e seus agrupamentos, identificamos artigos-chave (artigos seminais e artigos relevantes) que suportam as diferentes áreas de Avaliação de Pesquisa. O principal resultado é o mapa de citações organizado em seis grupos (clusters), fornece uma visão geral rápida desses territórios interligados, dando uma compreensão dessas diferentes áreas de conhecimento. Esta é uma imagem clara deste tópico, útil para acadêmicos, pesquisadores juniores, bem como dos diversos interessados práticos com interesse na política de avaliação de pesquisa ou na avaliação da qualidade da pesquisa e sua implementação. Identificamos publicações seminais e publicações relevantes com elevados índices de citação. Este estudo é um sólido ponto de partida para uma futura e profunda revisão da literatura sobre "avaliação de pesquisas".

Palavras-chave: Avaliação da Pesquisa. Estrutura intelectual. Análise bibliométrica. Visualização de dados.

Abstract: Research Evaluation topic has been the object of intense study both for its importance for multiple purposes, and as for theoretical development model system on a continuous increase of quality of research performance and societal and innovation impact. The objective of this exploratory study is to capture and mapping intellectual structure of Research Evaluation. Using networks citations we capture the latent building blocks and drivers of this theme. Data visualization and visual analytics were helpful to make an exploration of literature background; by observing citation networks and its clusters we identify key papers (seminal and highest papers) that support different areas of Research Evaluation. Main result is the Citation map with six clusters, where each cluster gives us a quick overview of those interlinked territories that provide an understanding of those different knowledge areas. This is a clear picture of this topic, useful for academic scholars, novice researchers and the practical stakeholders, interest in research evaluation policy or assessing the quality of research and its implementation. We identify Seminal and Highest citation publications. This study is the solid starting point for a deep literature review on "research evaluation".

Keywords: Research evaluation. Intellectual structure. Bibliometric analysis. Data visualization.

1 Introduction

Carol WEISS defines evaluation as "the systematic assessment of the operation and/or the outcomes of a program or policy, compared to a set of explicit or implicit standards, as a means of contributing to the improvement of the program or policy" (WEISS, 1998, p. 4). At science contexts evaluation is a natural issue as a mechanism of certification and control of research quality.

Traditionally Research Evaluation had been done by peers to evaluate individual researchers, with a qualitative approach. The use of scientific quantitative indicators emerged from 1950-1970 years. The development of Information and Communication Systems and the emergence of the Internet facilitated the access to global scientific literature and also publications bibliometric (citations, networks citations and other indicators at macro, meso and micro levels of analysis). Those contextual conditions have facilitated the emergence of a new quantitative evaluation stage.

Eugene Garfield is a bibliometric key author; he conceptualized the Science Citation concept, and its related Science Citation Index (GARFIELD, 1955). Additionally, at a practical level, Garfield founded the Institute for Scientific Information (ISI) in the 1960s (now Clarivate Analytics, see clarivate.com).

Thus the bibliometric stage has its origins as early as the beginning of the last century, but, it became data-driven in 1955 with the introduction of the science citation index. Another mark document in this research quantitative evaluation stage is the Frascati Manual published in 1963, which defined some indicators for monitoring and comparing nation-states from a statistical approach (OECD, 1962). Bibliometric is the "statistical analysis of books, articles, or other publications" (OECD, 2002, p. 203) and provides useful performance overview.

Scientometrics is a specialized field developed from the 1960s that support the study of science, technology, and innovation from a quantitative perspective (LEYDESDORFF; MILOJEVIĆ, 2015). Scientometrics was first defined by Nalimov and Mulcjenko (NALIMOV; MULCJENKO, p. 2) as developing "the quantitative methods of the research on the development of science as an informational process". The landscape of Scientometrics is a territory that results from the intersection of Sociology of Science, Science Policy and Information Science (Figure 1). Those knowledge areas are simultaneously users and suppliers of indicators that evolve from the tensions and controversies fed by their various actors.

Figure 1 - Landscape of Scientometrics
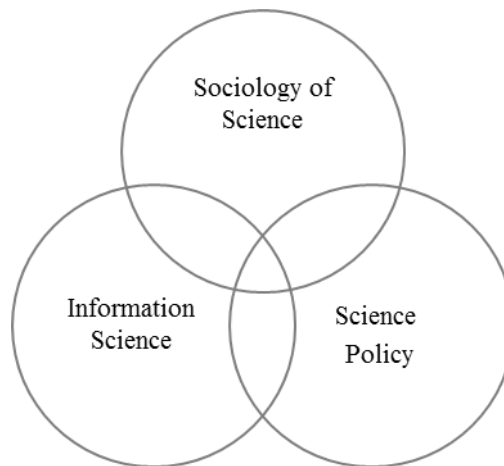
Source: Based on Leydesdorff and Milojević (2015)

Citation expresses the scholarly activity and helps to understand how this activity develops by the behaviour publication and citation of its authors. Looking to citation networks help the understanding of how one field or subfield of research is structured and develops, as Garfield expressed: "If the literature of science reflects the activities of science, a comprehensive, multidisciplinary citation index can provide an interesting view of these activities. This view can shed some useful light on both the structure of science and the process of scientific development" (GARFIELD, 1979, p. 62).

We know that issue of citations among scientists is controversial. Being cited shows impact, build reputation and citations are part of a reward system of science. Ther misuse of citations sometimes are criticized because citations not reflecte individual scientific contribution (BAIRD; OPPENHEIM, 1994; AKSNES, 2003; AKSNES; RIP, 2009). On the other hand, using networks citations at macro level can give a useful broad picture of the scientific structure.

The importance of the research evaluation as a public policy issue at a global scale has been established and has been a significant interest in the research community to evaluate the research activities through the use of the scientometric method.

Research evaluation can be used for multiple purposes: to provide accountability; for analysis and learning; for funding allocation; for advocacy and networking (PORTER *et al.*, 2007; GARNER *et al.*, 2012; GUTHRIE *et al.*, 2013; PINHO; ROSA, 2016).

Research evaluation is an umbrella concept that crosses diverse micro, meso and macro scales for many purposes and performed with different approaches. This

complexity can bring fragmented and misuse of the concept. This problem calls for a clear mapping territory concept.

This paper aims to explore the intellectual background structure of "research evaluation" topic. With the expansion of Research Evaluation practices at organization, institution and country levels, some research theories, topics and results have been published within this interdisciplinary theme.

The object of study is Research Evaluation. Because research is a relevant activity that creates new knowledge, products and processes there is a need to evaluate this activity. We aim to understand the state of art of Research Evaluation topic and clarify this concept. Consequently, our main research question is: What is Research Evaluation?

To achieve the answer to this question we start by performing an exploratory study, by identifying some key blocks related to the structure of this topic by searching answers to the following specific questions:

a) What are the seminal, core, relevant and review documents on Research Evaluation topic?
b) What are the structural properties of scientific publication?
c) What is the intellectual latent structure that drives Research Evaluation?

2 Methods and material

In January 2018 a search on Web of Science ® (WoS) inside some of its databases (SCI-EXPANDED, SSCI, A&HCI) was performed in the field TOPIC with the terms: "research evaluation" *OR* "research assessment" *OR* "research measurement". The TOPIC includes searching in title, abstract, Author Keywords and Keywords Plus®. The terms were chosen because the use of these terms appears in a related way (JUBB, 2013; BARTOL *et al.*, 2014; GLANZEL, THIJS; DEBACKERE, 2014).

The period selected was 2006 to 2017, and the search resulted in 1,483 publications. Those publications were analyzed as follows: a) a description of the properties of our sample literature in terms of measures such as the number of articles on Research Evaluation topic, the most prolific journals, types of documents and the annual distribution of publication; b) the identification of the intellectual structure, divides into two steps: first, the decomposition of the sample literature into disciplinary and subject categories; second, a visualization of intellectual structure using network analysis. For this last stage, the sample of publications was input in CitNetExplorer.

We analyze that citation network and then we cluster the publications. We obtained six clusters; thus, to deeply investigate those clusters we then make cluster analysis and we obtained the subclusters. The clustering technique used by CitNetExplorer is discussed in Waltman and Van Eck (2012). The resolution parameter determined the level of detail at which cluster. The higher the value of the parameter, the larger the number of clusters that will be obtained (VAN ECK; WALTMAN, 2014). In this study, for all cluster analysis, this parameter was set to 0.75.
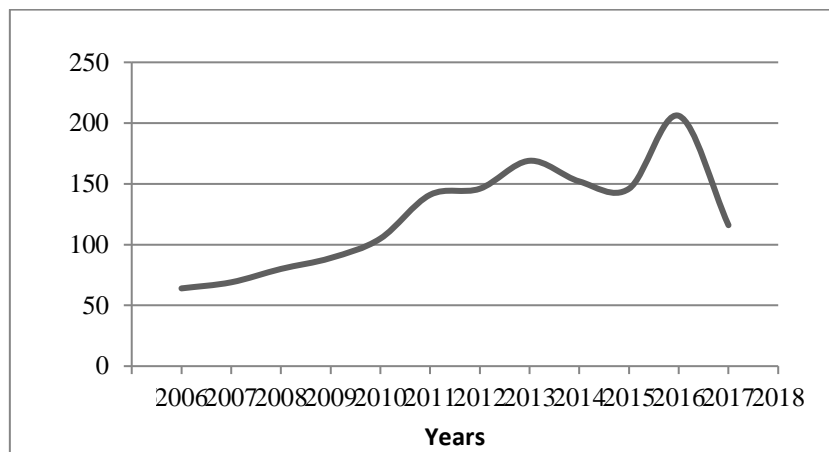
3 Descriptive analysis

In this section, we present some results related to document type, the most prolific journals, the evolution publication years and the most prolific subject areas.

The majority of documents were articles (83.2%) followed by reviews (7.1%), proceedings papers (3.5%), and others (6.1%). In this sample, the most prolific journals were: Scientometrics (171 articles); Research Evaluation (77); Journal of Informetrics (68); Plos One (22); Journal of International Business Studies (20); Journal of the Association for Information Science and Technology (20); Research policy (18); Higher education (16); Journal of the American Society for Information Science and Technology (16); Profesional de la informacion (14); Qualitative Health Research (13); Studies in Higher Education (11).

Figure 2 shows the number of publications per publication year, during the 2006-2017 period. The average growth of publications, between 2006 and 2016, was 13.4%. Considering these 10 years (2006-2016) we can define 3 periods. First one, from 2006 to 2010, has slow growth in publications, from 64 publications in 2006 to 105 in 2010. The second period begins in 2011 and finishes in 2014 has a stronger growth, with a publication average around the 150 articles. The third period (2015-2016) begins with the year 2015 with a slight slowdown, but the year 2016 reaches 206 articles. Notice that the most remarkable year for the annual growth rate is 2016 with 41%, followed by 2001 with 34%. These data provide further evidence that the research evaluation theme has been a mature research area since 2011.

Figure 2 - Annual distribution of publications
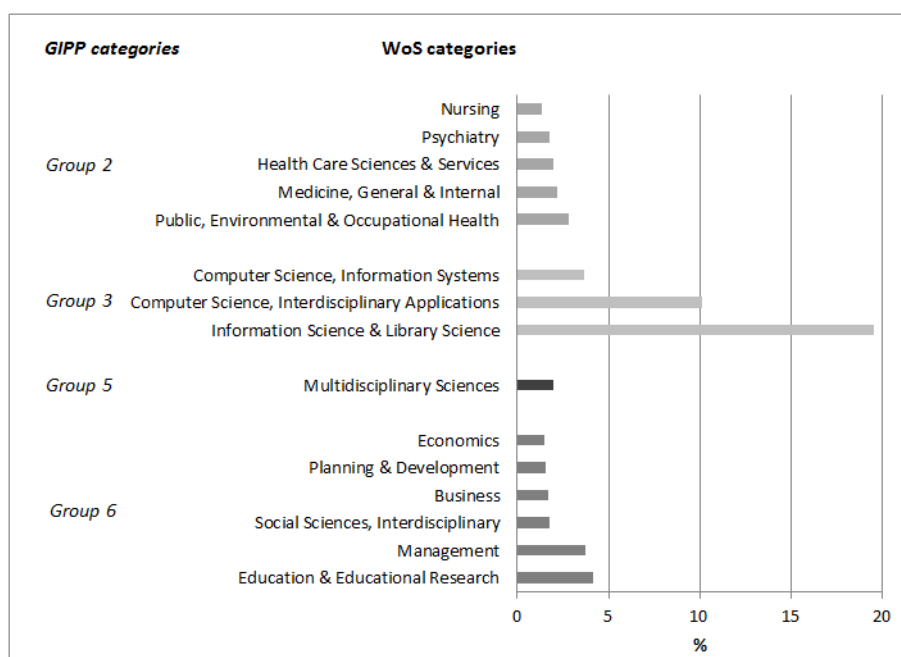
3.1 Intellectual structure

To discover the intellectual structure we decompose of the sample literature into disciplinary and subject categories, and then we create and analyze a citation network of publications and conducted a cluster analysis.

3.2 Disciplinary and subject categories

The decomposition of scientific literature into disciplinary and subject categories can map the contribution and interest of the various disciplines for the development of the theme. To analyze the composition of research evaluation literature, we use the WoS scheme (252 subject categories) and GIPP scheme (6 categories). The WoS categorization assigning each journal to one or more subject categories (OSWALD, 2007). Consequently, for our sample of 1,483 publications we have obtained a higher frequency (2,413 categories), that is an article can appear associated with more than one category (overlapping of subject categories). The GIPP schema is based on an aggregation of the WoS subject categories into six broad disciplines: 1-Arts & Humanities; 2-Clinical, Pre-Clinical & Health; 3-Engineering & Technology; 4-Life Sciences; 5-Physical Sciences; 6-Social Sciences (OSWALD, 2007). As Figure 3 shows, the top 15 subject categories and the correspondent GIPP categories. The two categories with highest frequencies, Information Science & Library Science (20%) and Computer Science, Interdisciplinary Applications (10%), are associated with Engineering and Technology GIPP group. The following categories with higher frequencies are related to the Social Sciences GIPP group, specifically Education & Educational Research (4%) and Management (4%).

From those results, it is clear that the topic of research evaluation is supported by a broad spectrum of disciplines that contribute to its theoretical development and practical implementation. This interdisciplinary nature provides the energy of the evolution of its background and broad scientific interest.

Figure 3 – Top 15 subject categories of Web of Science (WoS)

## 3.3 Network visualization and cluster analysis with CitnetExplorer

Science can be seen as an evolving network system. Networks are a powerful language to describe the patterns of interactions that build complex systems (NEWMAN, 2001). The structural properties of scientific publication can be explicited by citation networks.

We use CitNetExplorer software to obtain a citation network consisting of 2,243 publications and 9,011 citations links. Note that in this network there are 760 additional publications apart to 1,483 the publications that constitute our initial sample. These are publications that are cited by at least in five publications included in the sample. For more details see VAN ECK; WALTMAN (2014). In this paper, the authors present this software tool for analyzing and visualizing citation networks of scientific publications.

Timeline-based network visualization of the 70 most-cited publications is shown in Figure 4. The location of a publication in the vertical dimension is determined by the

year in which the publication appeared, and in the horizontal dimension, by the closeness of publications in the citation networks (VAN ECK; WALTMAN, 2014).

Figure 4 - Citation network and its clusters
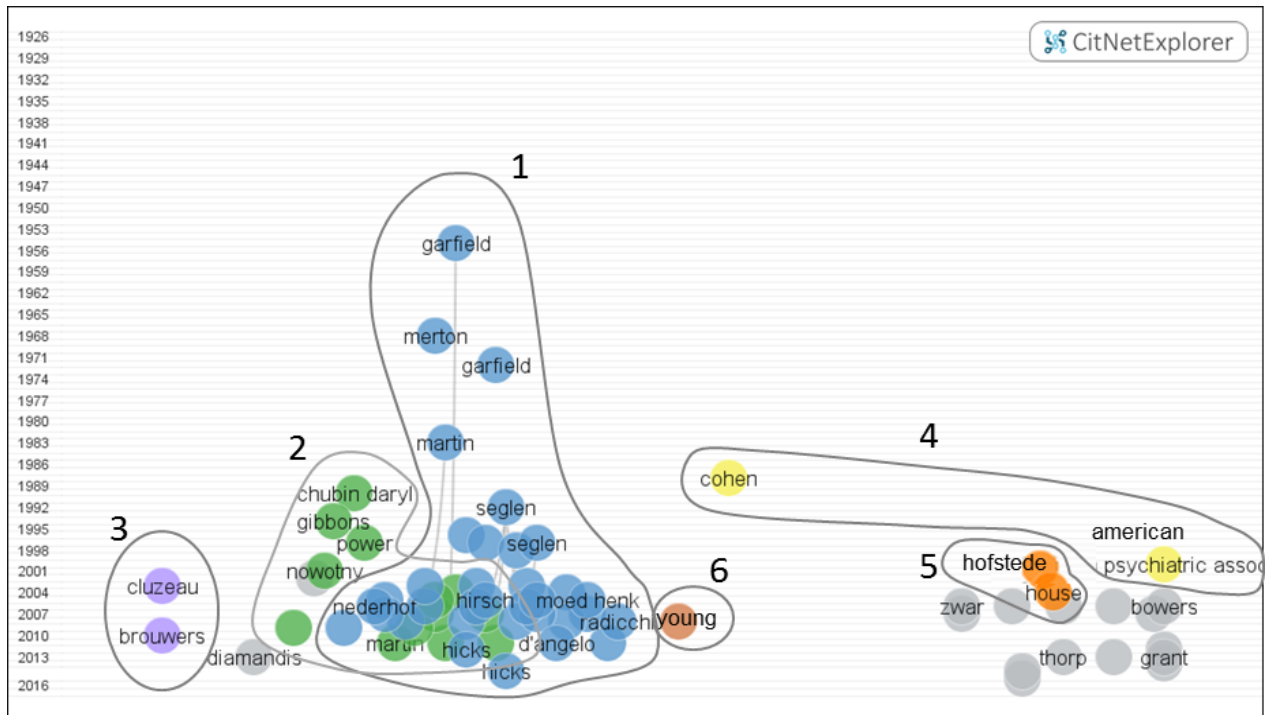
Source: Authors

The oldest paper of this citation network came from chemist Alfred Lotka with the title "The Frequency Distribution of Scientific Productivity" (LOTKA, 1926). Note this paper is not visible in Figure 4 because do not belong to the 70 most cited publications. In this paper, Alfred Lotka examined the frequency distribution of scientific productivity of chemists and physicists, in chemical abstracts 1907–1916, and he noted that the number of persons making n contributions is about $1/n^2$ of those making one, and the proportion of all contributors that make a single contribution is about 60%. This logarithmic, the "Lotka's law", is a relevant example of the contribution from Statistic discipline to understand a social phenomenon: the scientific collaboration explained by the publication in co-authorship.

To identify specific "research territories" in Research Evaluation, a cluster analysis was performed in CitnetExplorer; six subgroups were identified (Table 1 and Figure 4). Each cluster has publications that are strongly connected, in terms of citation networks (VAN ECK; WALTMAN, 2014). Note that 534 publications do not belong to any cluster.

Table 1 – Citation network information for the six clusters

Cluster	Number of publications [a]	Number of publications with 10 or more citations	Number of citations
1	1083	221	6385
2	409	38	409
3	78	3	78
4	69	4	69
5	68	2	68
6	28	1	61

[a] number of citations of the publication within the citation network being analyzed and obtained from CitNetExplorer

Source: Authors

The clusters 1 and 2 have the greatest number of publications, the total number of citations and number of publications with 10 or more citations (Table 1). As is seen in Figure 4 the links of these two clusters (1 and 2) are closely related to each other. These two clusters are the most relevant to identify specific research areas in Research Evaluation.

3.3.1 Cluster 1

Cluster 1 seems to be the core of the research evaluation theme. From this trunk, the other clusters are evolving and affirming their identity. Considering only the 70 most cited publications, the oldest goes back to 1955 with the paper of Eugene Garfield publish in *Science* (1955). In this article, he suggested that there is a need to create a database that would allow access to cited articles, i. e., for each article, must be available the cited articles. Based on this link between the citing article and cited article, Eugene Garfield proposed the Science Citation Index (SCI). This concept of citation indexing begins as a measure of information retrieval and has evolved to turns a tool for research evaluation (GARFIELD, 1972; 1996). Thus, "SCI's multidisciplinary database has two purposes: first, to identify what each scientist has published, and second, where and how often the papers by that scientist are cited. Hence, the SCI has always been divided into two author-based parts: the Source Author Index and the Citation Index. By extension, one can also determine what each institution and country has published and how often their papers are cited. The WoS - the SCI's electronic version- links these two functions: an author's publication can be listed by chronology, by the journal, or by citation frequency. It also allows searching for scientists who have published over a given period of years" (GARFIELD, 2007).

Another seminal document of cluster 1, comes from Sociology of Science discipline the Merton book (1967) "On Theoretical Sociology. Five Essays, Old and New". This book supports a debate about scholarly work that includes publication and citation behaviour.

With the 1083 publications of cluster 1, a cluster analysis was performed and results in 5 subclusters (Table 2 and Figure 5).
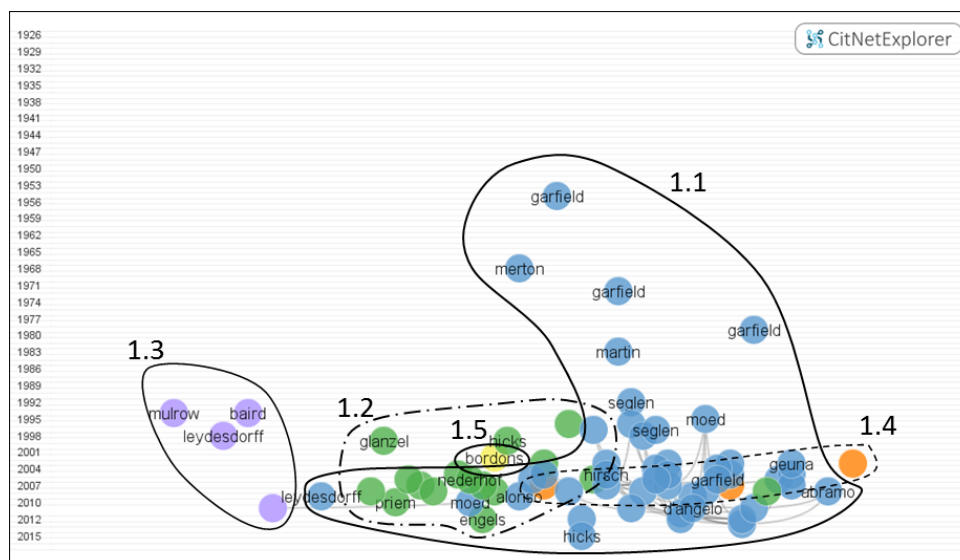
Table 2 – Principal publications from subclusters of Cluster 1

		Author (year)	Number of citations[a]	Title
Subcluster 1.1	Seminal	Garfield (1955)	28	Citation indexes for science: a new dimension in documentation through association of ideas.
		Garfield (1972)	17	Citation Analysis As a Tool in Journal Evaluation - Journals Can be Ranked by Frequency and Impact of Citations for Science Policy Studies.
		Merton (1967)	22	*On Theoretical Sociology. Five Essays, Old and New.*
	Highest citation	Hirsch (2005)	134	An index to quantify an individual's scientific research output.
		Moed (2005)	73	*Citation Analysis in Research Evaluation.*
		Van Raan (2005)	43	Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods.
		Hicks (2012)	34	Performance-based university research funding systems.
	Seminal	Macroberts and Macroberts (1989)	15	Problems of citation analysis: A critical review.
	Highest citation	Nederhof *et al.* (1993)	41	Research performance indicators for university departments: A study of an agricultural university.
		Meho and Yang (2007)	38	Impact of data sources on citation counts and rankings of LIS faculty: *Web of science* versus *Scopus* and *Google scholar.*
Subcl. 1.3	Seminal	Garfield (1986)	6	Mapping cholera research and the impact of Sambhu Nath De of Calcutta
	Highest citation	Baird and Oppenheim (1994)	17	Do citations matter?
		Leydesdorff (1998)	17	Theories of citation?
Subcl. 1.4	Seminal	Liebowitz and Palmer (1984)	8	Assessing the relative impacts of economics journals.
	Highest citation	Oswald (2007)	16	An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-Makers.
Subcl. 1.5	Seminal & Highest citation	Bordons *et al.* (2002)	12	Advantages and limitations in the use of impact factor measures for the assessment of research performance.

[a] the number of citations of the publication within the citation network being analyzed and obtained from CitNetExplorer

Source: Authors

Figure 5 – Citation network of cluster 1 and its subclusters

The subcluster 1.1 includes 705 publications. Notice that the seminal publications (GARFIELD, 1955; MERTON, 1967; GARFIELD, 1972) above mentioned belonging to this subcluster 1.1 Besides these publications, it seems that four papers (with the highest citation) have a driver role on the development of the theme of Research Evaluation mainly on quantitative indicators, such as Hirsch (2005), Moed (2005), Van Raan (2005) and Hicks (2012).

In 2005 Jorge Hirsch designed a new evaluation index (h-index) to characterize the scientific output of a researcher by measuring the impact of the scientist's publications only in terms of the received citations (HIRSCH, 2005). In same year Henk F. Moed publishes the book "Citation Analysis in Research Evaluation" where he advocates the use of citation analysis in social sciences and humanities and he notes that research quality is not merely a social construct but it relates to a quality intrinsic to the research itself (MOED, 2005). He also draws attention that this concept cannot be defined and measured in Social Sciences and Humanities in the same way as in other science. This position led him to consider the diversity dimension as an integral part of the evaluation design (diversity in publication sources and diversity in publication languages).

Anthony Van Raan also publishes in 2005 a relevant article where he describes the conceptual and methodological problems of misleading bibliometric methods use which can cause damage to universities, institutes and individual scientists (VAN RAAN, 2005). Diana Hicks focuses on university research evaluation and related university research funding systems (HICKS, 2009; 2012; 2017).

The subcluster 1.2 contains 254 publications. The article published in 1989 present a critical review of the problems related to citation analysis, warning users of citation-based literature that they should proceed cautiously (MACROBERTS; MACROBERTS, 1989). Within this subcluster, the most cited article is related to research performance indicators for the university, at a departmental level, in the natural and life sciences, the social and behavioural sciences, and the humanities (NEDERHOF *et al.*, 1993). Another article with many citations is about the impact of using different citation databases (*WoS, Scopus*® and *Google Scholar*®) on the citation counts and rankings of scholars; additionally, the authors recommend the use of more than one data source to have a more accurate and comprehensive picture of the scholarly impact of authors (MEHO; YANG, 2007).

Subcluster 1.3 contains 65 publications. In 1986, Garfield used an historiograph, a graphical representation that helps the visualization of the historical development of research based on the most highly cited papers, both within the field and in all of science, for mapping cholera research (GARFIELD, 1986). In this seminal paper Garfield proposed the concept of delayed recognition, and concluded that De's 1959 paper in *Nature*: "while initially unrecognized, today is considered a milestone in the history of cholera research." and "great breakthroughs in science are often initially overlooked or simply ignored, as in De's discovery of the existence of a cholera enterotoxin" (GARFIELD, 1986).

The article entitled "Do citations matter?" have the highest citation in the subcluster 1.3. In this paper, Laura Baird and Charles Oppenheim analyze some of the criticisms of citation counting and conclude that citation studies remain a valid method of analysis of individuals', institutions', or journals' impact, but need to be used with caution and in conjunction with other measures (BAIRD; OPPENHEIM, 1994).

In other highest citation article, Leydesdorff (1998) draws attention to the use of Geometrical representations ('mappings') as a reach source for sociological interpretation of citations and citation practices.

Subcluster 1.4 includes 49 publications. In 1984, Liebowitz and Palmer publish an article in the *Journal of Economic Literature*. They propose a ranking of journals based on citations (LIEBOWITZ; PALMER, 1984). Later, those authors propose to improve impact factor of journals by (1) standardizing journals to compensate for size and age differentials; (2) including a much larger number of journals; (3) using an iterative process

to "impact adjust" the number of citations received by individual journals (COATS, 1971; VOOS; DAGAEV, 1976).

The most cited article in the subcluster 1.4 is the critical one written by Andrew Oswald (2007). The author questioned the use of journal rankings as a measure of "research quality". He examines data on citations in articles published during 25 years and he advocates that it is better to write the best article published in an issue of a medium-quality journal than four articles published in an issue of an elite journal. Decision-makers need to understand what is research quality (OSWALD, 2007, p. 21).

Maria Bordons, Fernandez and Gomez (2002) paper is the seminar one of subcluster 1.5. In this paper, the authors present the usefulness of the impact factor measures in macro, meso and micro analyse, at Spanish production context. Some main advantages of using impact factor (IF), such as the great accessibility of impact factor and its ready-to-use nature are pointed out.

Additionally, they point several limitations such as:

– A fixed set of journals could be convenient for international analysis. Very few journals from peripheral countries are covered in SCI and SSCI. However, when domestic journals from these countries are covered, a decrease in the average IF of the country production is produced, since national journals usually show very low impact factors. This could influence time series data and should be taken into account by analysts.

– Calculation of IF for domestic journals not covered by ISI databases may be a useful way of complementing data from ISI. However, this is a very laborious, expensive and time-consuming task.

– The widespread use of IF measures within the research evaluation process have produced some negative effects. Abuse and incorrect use of IF measures are the underlying reasons. The priority of international vs. national journals in the agenda of scientists (with the corresponding impoverishment of the domestic journals); priority of international vs. national research subjects (particularly in peripheral countries) and a lesser consideration of "slow" evolving disciplines (in which low impact factors are the norm) are some of the consequences observed. Impact factor users should be aware of these effects and make sensible use of impact factor indicator.

3.3.2 Cluster 2

The cluster 2 has 409 publications and we found six subclusters (Table 3 and Figure 6). The subcluster 2.1 contains 145 publications and the main seminal article is the
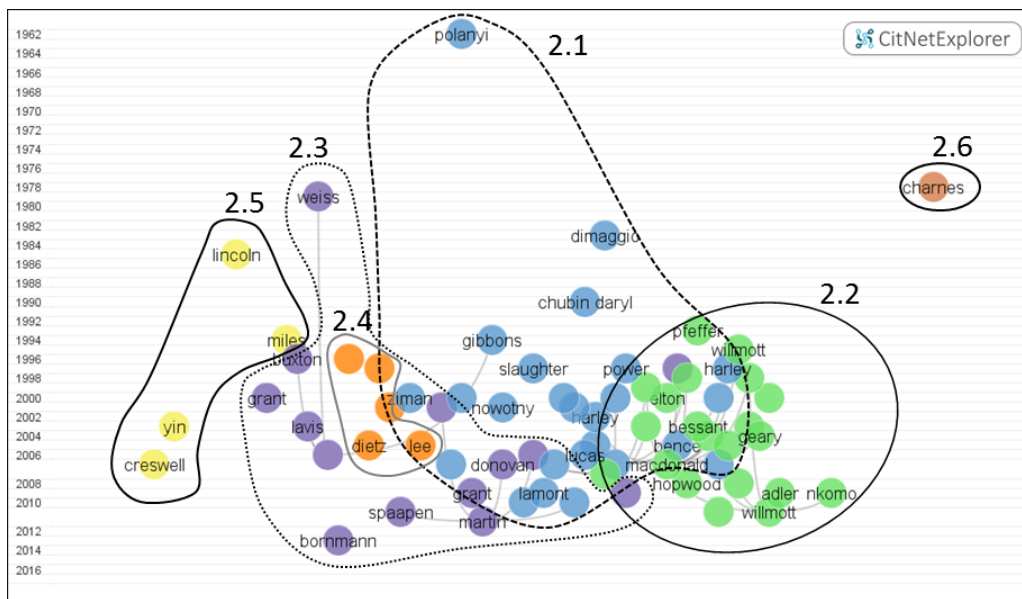
publication of Michael Polanyi in 1962 that reflect on the myth of isolated work of the scientists, that choose its research problems and publish alone (POLANYI, 1962). Instead, scientific activities must be coordinated. It is dynamic coordination with mutual adjustment of independent initiatives within the same system.

Table 3 - Principal publications from subclusters of Cluster 2
Source: Authors

		Author (year)	Number of citations[a]	Title
Subcluster 2.1	Seminal	Polanyi (1962)	7	The Republic of science: Its political and economic theory.
		Glaser and Strauss (1967)	5	*The Discovery of Grounded Theory: Strategies for Qualitative Research.*
		Kuhn (1962)	5	*The structure of scientific revolutions.*
		Dimaggio and Powell (1983)	7	The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields.
	Highest citation	Gibbons *et al.* (1994)	23	The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies
		Nowotny, Scott and Gibbons (2001)	15	*Re-thinking science: Knowledge and the public in an age of uncertainty.*
Subcluster 2.2	Seminal	Mahoney (1977)	4	Publication prejudices: An experimental study of confirmatory bias in the peer review system.
		Ceci and Peters (1982)	4	Peer Review: A Study of Reliability.
	Highest citation	Macdonald and Kam (2007)	16	Ring a ring o'roses: Quality journals and gamesmanship in management studies.
		Geary, Marriott and Rowlinson (2004)	14	Journal Rankings in Business and Management and the 2001 Research Assessment Exercise in the UK.
		Starbuck (2005)	14	How Much Better are the Most-Prestigious Journals? The Statistics of Academic Publication.
Subcluster 2.3	Seminal	WEISS (1979)	6	The many meanings of research utilization.
	Highest citation	Martin (2011)	14	The Research Excellence Framework and the 'impact agenda': are we creating a Frankenstein monster?
		Buxton and Hanney (1996)	11	How Can Payback from Health Services Research Be Assessed?
		Donovan (2007)	10	The qualitative future of research evaluation.
Subcluster 2.4	Seminal	Levin and Stephan (1991)	2	Research productivity over the life cycle: Evidence for academic scientists.
	Highest citation	Dietz and Bozeman (2005)	7	Academic careers, patents, and productivity: industry experience as scientific and technical human capital.
		Lee and Bozeman (2005)	7	The Impact of Research Collaboration on Scientific Productivity.
Subcl. 2.5	Seminal & Highest citation	Lincoln and Guba (1985)	6	*Naturalistic Inquiry.*
		Miles *et al.* (2013)	6	*Qualitative Data Analysis.*
Subcl. 2.6	Seminal & Highest citation	Charnes *et al.* (1978)	5	Measuring the efficiency of decision making units.

[a] the number of citations of the publication within the citation network being analyzed and obtained from CitNetExplorer

Figure 6 - Citation network of cluster 2 and its subclusters

Follow this idea of knowledge production as a result of the coordination of a social space where tacit and explicit knowledge interplay, Paul DiMaggio and Walter Powell take an organizational approach conveyed by the article published in 1983. They defend that the collective rationality occurs at the organizational context where the mechanisms of isomorphic can change (DIMAGGIO; POWELL, 1983).

Another seminal work is the book of Thomas S. Kuhn "The structure of scientific revolutions" that explores the psychology of belief that governs the acceptance of new concepts and innovations in science. Kuhn showed that the history of science is not one of linear, arguing that transformative ideas come from radical shifts of vision in which a multitude of non-rational and non-empirical factors come into play. He also points out that the revolutions in science are breakthrough moments that disrupt accepted thinking and this causes a change in the way of doing science and applying it. Kuhn used the word "paradigm" to describe this conceptual aggregation of theories, methods, and assumptions about reality that guide researchers a global platform to retrieve data, elaborate theories, and solve problems (KUHN, 1962).

Another literature pillar of this Subcluster 2.1 is the book entitled "The Discovery of Grounded Theory: Strategies for Qualitative Research" from Glaser and Strauss (1967). The book lays out a complete system of building theories from qualitative data and defined rigour in qualitative research.

The highest citation in this subcluster 2.1 is the book "The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies". The main idea is that knowledge production was changed from Mode1, where knowledge is generated within a disciplinary, primarily cognitive, context, to Mode2 where knowledge is created in broader, transdisciplinarity social and economic with a context of the application (GIBBONS *et al.*, 1994).

The second most cited reference is the book of Nowotny, Scott, and Gibbons (2001) "Re-thinking science: Knowledge and the public in an age of uncertainty". The authors argue that changes inside society make an impact on science not only in its research practices but also deep in its epistemological core. This co-evolution requires a complete re-thinking on which a new social contract between science and society might be constructed.

The subcluster 2.2 enclose 123 publications. The seminal article of cluster 2.2 is from Michael Mahoney (1977). The author asked 75 journal reviewers about the peer review system, in clinical research. He found that reviewers were strongly biased against manuscripts which reported results contrary to their theoretical perspective. Another seminal paper (CECI; PETERS, 1982) focus on the reliability of peer review system analyzing bias in peer reviews identified several kinds of bias such as *affiliation* bias, meaning that researchers from prominent institutions are favoured in peer review. This is an important issue when jobs and salaries of academics "often depend on the reviews of their colleagues" (CECI; PETERS, 1982, p. 44).

The most cited articles in subcluster 2.2 focus on the quality journal. Stuart Macdonald and Jacqueline Kam focus on the pressure to publish in quality journals and the related behaviour game. First, they started to criticize the fact of "a paper in one of the quality journals of Management Studies is much more important as a unit of measurement than as a contribution to knowledge" (MACDONALD; KAM, 2007, p. 640) from the perspective of academic performance and related academic funding. Second, they put the question: what is a quality journal? Then they analyze the cost of playing the game in this competitive market, played by universities, departments, and authors. One of the problems of this game is the impact of "rewards for publishing attach to the content of papers, to what is published rather than where it is published" (2007, p. 640).

In 2004, Geary, Marriott, and Rowlinson make a journal rankings in Business and Management based on the 2001 Research Assessment Exercise in the UK (GEARY; MARRIOTT; ROWLINSON, 2004). William Starbuck, in 2005, questioned about the

contribution of the articles published in high-prestige journals to knowledge? The author's article uses a statistical theory of review processes to draw inferences about differences value between articles in more-prestigious versus less-prestigious journals (STARBUCK, 2005).

The subcluster 2.3 contains 80 publications and the seminal paper is from WEISS (1979). Carol WEISS presents a literature review about research utilization. She extracted seven different meanings associated with this concept: 1) knowledge-driven; 2) problem-solving; 3): interactive; 4) political; 5) tactical; 6) enlightenment and 7) social science.

The three papers with the highest citation in subcluster 2.3 are Martin (2011), Buxton and Hanney (1996), and Donovan (2007). Ben Martin examines the origins of research evaluation, particularly in the UK. From an accountability perspective, the mechanisms for assessing research performance have become more sophisticated and costly. In the United Kingdom, the Research Assessment Exercise (RAE) evolved from an initially simple framework to something much more complex and onerous. As the RAE now gives way to the Research Excellence Framework (REF), and an 'impact assessment' is being added to the process and this is a complex indicator. Notice that the "central and inescapable problem with any assessment system, namely that once you measure a system, you irrevocably change it" (MARTIN, 2011, p. 250).

Buxton and Hanney applied the concept of payback at health research and development and they identified five main categories of payback: 1) KP (knowledge production); 2) RTCB (research targeting and capacity building); 3) IPPD (informing policy and product development); 4) HB (health and health sector benefits); 5) BEB (broader economic benefits). They developed a new conceptual model of how and where payback may occur (BUXTON; HANNEY, 1996). The model "characterizes research projects in terms of Inputs, Processes, and Primary Outputs.

Claire Donovan considered two types of evaluation strategies: 1) qualitative approaches, and 2) quantitative approaches (DONOVAN, 2007). The traditional qualitative approach is based on peer review. With recent convergence interest of some research fields (sociology, statistics, and information) and with the available data in bibliographical platforms (*WoS®, Scopus®* or *Scielo®*) quantitative approach has increased with a focus on scientific articles. She defends a qualitative future of research evaluation because "advances in the use of quality and impact metrics have followed a trajectory away from the unreflexive use of standardized quantitative metrics divorced from expert peer interpretation, towards triangulation of quantitative data, contextual

analysis and placing a renewed and greater value on peer judgment combined with stakeholder perspectives" (DONOVAN, 2007, p. 594).

The seminal article (LEVIN; STEPHAN, 1991) of subcluster 2.4 (28 publications) focus on the relationship between the research productivity of PhD scientists and age. They used a computer algorithm to link journal-publication data contained in the Science Citation Index (SCI) with the Survey of Doctorate Recipients (SDR), in four fields, at USA context (LEVIN; STEPHAN, 1991).

Two articles published in 2005 have the highest citation in subcluster 2.4. Dietz and Bozeman examined career patterns within the industrial, academic, and governmental sectors and their relation to the publication and patent productivity of scientists and engineers working at university-based research centres, in the USA. They hypothesized "that among university scientists, intersectoral changes in jobs throughout the career provide access to new social networks and scientific and technical human capital, which will result in higher productivity" (DIETZ; BOZEMAN, 2005, p. 353). Their research question is "what effects do job transformations and career patterns have on productivity (as measured in publication and patent counts) over the career life cycle?" and the central hypothesis to be tested is "among university scientists, intersectoral changes in jobs throughout the career will provide access to new social networks, resulting in higher productivity, as measured in publications and patents" (DIETZ; BOZEMAN, 2005, p. 353).

Lee and Bozeman examine the assumption that research collaboration has a positive effect on publishing productivity. They focus at the individual level but they recognize "the most important benefits of collaboration may accrue to groups, institutions, and scientific fields" (LEE; BOZEMAN, 2005, p. 349).

Subcluster 2.5 contains 25 publications. The seminal book "Naturalistic Inquiry" of Lincoln and Guba, publish in 1985, propose an alternative paradigm a "naturalistic" rather than "rationalistic" method of inquiry, in which the researcher avoids manipulating research outcomes. This book can help all social scientists involved with questions of qualitative and quantitative methodology (LINCOLN; GUBA, 1985)

The third edition of Miles, Huberman and Saldaña's book "Qualitative Data Analysis" have the highest citation of subcluster 2.5 (MILES; HUBERMAN; SALDAÑA, 2013). This book presents the fundamentals of research design and data management, by describing five distinct methods of analysis: exploring, describing, ordering, explaining, and predicting.

The seminal and highest citation of the subcluster 2.6 (8 publications) is the article of Charnes, Cooper, and Rhodes that provide a nonlinear (nonconvex) programming model based on connections between engineering and economic approaches to efficiency evaluate and control managerial behaviour in public programs (CHARNES; COOPER; RHODES, 1978).

Table 4 - Principal publications of Clusters 3, 4, 5 and 6

		Author (year)	Number of citations[a]	Title
Cluster 3	Seminal	Woolf *et al.* (1999)	7	Clinical guidelines: potential benefits limitations and harms of clinical guidelines.
	Highest citation	Brouwers *et al.* (2010)	18	AGREE II: advancing guideline development, reporting and evaluation in health care.
		Cluzeau (2003)	16	Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project.
Cluster 4	Seminal	Kuhn (1962)	9	*The structure of scientific revolutions.*
		Beck *et al.* (1961)	7	An inventory for measuring depression.
	Highest citation	APA (1994)	24	*Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*
		Cohen (1988)	12	*Statistical Power Analysis for the Behavioral Sciences.*
		Landis and Koch (1977)	10	The Measurement of Observer Agreement for Categorical Data.
Cluster 5	Seminal	Freeman (1979)	6	Centrality in Social Networks Conceptual Clarification.
		Anderson (1998)	6	Structural equation modeling in practice: A review and recommended two-step approach.
		Cronbach and Meehl (1955)	5	Construct Validity in Psychological Tests.
	Highest citation	House *et al.* (2004)	13	*Culture, leadership, and organizations: The GLOBE study of 62 societies.*
		Hofstede (2001)	11	*Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations.*
		Hofstede (1980)	9	*Culture's consequences: International differences in work-related value.*
Cluster 6	Seminal	Callaham, Wears and Weber (2002)	6	Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals.
		Vinkler (2003)	5	Relations of relative scientometric indicators.
	Highest citation	Young, Ioannidis and Al-Ubaydli (2008)	10	Why Current Publication Practices May Distort Science.
		Patsopoulos, Analatos and Ioannidis (2005)	8	Relative citation impact of various study designs in the health sciences.

[a] the number of citations of the publication within the citation network being analyzed and obtained from CitNetExplorer

Source: Authors

3.3.3. Cluster 3

The cluster 3 (Table 4) with 78 publications is defined by evaluation in health care scope. The seminar paper of Steven Woolf, Richard Grol, Allen Hutchinson, Martin Eccles and Jeremy Grimshaw takes an organizational approach by defending the use of guidelines as a tool of quality framework evaluation, against which practice can be measured. This can help to make informed practice decisions and provide managers with a useful framework for assessing and managing the organization by supporting quality improvement activities (WOOLF *et al.*, 1999). Notice that the literature reviews, particularly systematic reviews, are considered relevant to integrate existing knowledge and detect knowledge gaps; initially, systematic research synthesis was applied to medicine and health and later replicated to other areas of research (ŠUBELJ; VAN ECK; WALTMAN, 2016). This rigorous procedure combined with consensus discussions among ordinary general practitioners and content experts can improve the timing and knowledge processes (LIEBOWITZ; PALMER, 1984; GARFIELD, 1986; VAN ECK; WALTMAN, 2015).

The most cited papers of cluster 3 are related to an international project, the AGREE Collaboration. This project has developed an international instrument for assessing the quality of the process and reporting of clinical practice guideline development that is reliable and is acceptable in European and non-European countries (CLUZEAU, 2003; BROUWERS *et al.*, 2010).

3.3.4 Cluster 4

The cluster 4 (69 publications) contains some articles from the perspective of the mechanism of science that can be studied by statistical mathematical methods. The seminal documents are from Kuhn (1962) and Beck *et al.* (1961). The seminal book of Kuhn (1962) is also seminal for cluster 2.1, so, the book was already mentioned.

Depression Inventory of Beck *et al.* (1961) is a 21-question multiple-choice self-report inventory, one of the most widely used psychometric tests for measuring the severity of depression self-report questionnaires and can be analyzed using techniques such as factor analysis. This instrument was originally developed to provide a quantitative assessment of the intensity of depression but this research instrument remains widely used in research in other fields (BECK *et al.*, 1961).

The papers with the highest citation in cluster 4 are from the American Psychiatric Association (1994), Cohen (1988) and Landis and Koch (1977).

Diagnostic and Statistical Manual of Mental Disorders (DSM) is published by the American Psychiatric Association (APA) and offers a common language and standard criteria for the classification of mental disorders. This categorial system was replicated to other fields of Research (APA, 1994).

Jacob Cohen provides a guide to power analysis in research planning that provides users of applied statistics with the tools they need for more effective analysis (COHEN, 1988). He gave his name to such measures as Cohen's kappa, Cohen's d, and Cohen's h.

The paper "The Measurement of Observer Agreement for Categorical Data" from Landis and Koch (1977) propose a "unified approach to the evaluation of observer agreement for categorical data by expressing the quantities which reflect the extent to which the observers agree among themselves as functions of observed proportions obtained from underlying multidimensional contingency tables. These functions are then used to produce test statistics for the relevant hypotheses concerning interobserver bias in the overall usage of the measurement scale and interobserver agreement on the classification of individual subjects. presents a general statistical methodology for the analysis of multivariate categorical data arising from observer reliability studies" (LANDIS; KOCH, 1977, p. 160).

3.3.5 Cluster 5

The seminal papers of cluster 5 (68 publications) are from Cronbach and Meehl (1955), Anderson (1998), and Freeman (1979). Lee Cronbach and Paul Meehl (1955) proposed that the development of a nomological net was essential to the measurement of a test's construct validity. Another contribution to study networks comes from James Anderson and David Gerbing, David with structural equation modelling in practice; this is a methodology for representing, estimating, and testing a network of relationships between measured variables and latent constructs (ANDERSON, 1998). The essay of Linton Freeman reviews the intuitive background for measures of structural centrality in social networks is reviewed and its existing measures are evaluated (FREEMAN, 1979).

The tree highest citation of cluster 5 are related to cultural dimensions at the organizational level. Hofstede's cultural dimensions theory is a framework for cross-cultural communication. This theory was one of the first quantifiable theories that could

be used to explain observed differences between cultures and has been widely used in several fields as a paradigm for research, particularly in cross-cultural psychology, international management, and cross-cultural communication (HOFSTEDE, 1980; 2001; HOUSE *et al.*, 2004).

3.3.6 Cluster 6

The seminal papers of cluster 6 (28 publications) are from Callaham, Wears and Weber (2002) and Vinkler (2003). Callaham and colleges (2002) observed how the journal prestige, publication bias, and other characteristics are associated with citation of published studies in peer-reviewed journals. Another seminal work of this cluster come from Péter Vinkler (2003) that defend the use of relative scientometric indicators for comparative evaluation of thematically different sets of journal papers.

The two highest citation articles in this cluster are from Young, Ioannidis and Al-Ubaydli (2008) and Patsopoulos, Analatos and Ioannidis (2005).

Neal Young and colleagues take a critical approach of the current system of publication in biomedical research that distorted the view of the reality of scientific data that are generated in the laboratory and clinic and this context can lead to misallocation of resources. This system must consider society's expectations and also scientists goals (YOUNG; IOANNIDIS; AL-UBAYDLI, 2008).

Patsopoulos and colleagues explore the relations between citation impact and the various study designs in the health sciences. They note that meta-analysis was cited significantly more often than other design. They concluded that "a citation does not guarantee the respect of the citing investigators. Occasionally a study may be cited only to be criticized or dismissed. Nevertheless, citation still means that the study is active in the scientific debate." (2005, p. 2366).

4 Discussion and conclusion

Exploratory analysis of citation patterns in "research evaluation" literature resulted in a solid theoretical background on this theme. The evolution stage and the consolidated stage were captured by analyzing the six clusters of publications.

The qualitative assessment of the citation networks serves to identify the characteristics of those cluster related to their seminal and most cited documents. Some tipping points were identified by linking some core publications.

An important finding from this result is that the research on the research evaluation theme is the interdisciplinary nature of its background with a broad scientific interest and practice application. Research Evaluation can be related to performance as a measurable result. Another main idea is related to research evaluation implementation; despite the controversies that research evaluation entails, there is a growing need for changes that call for the wise use of models, tools, and indicators. The use of performance-based research can bring positive consequences (accountability and transparency) but there is a need to combine quantitative and qualitative participative approaches to avoid perverse impacts.

Some tipping points were identified such as: a) in 1978, the journal "Scientometrics" was launched as a new medium for a new field of hard social science, the quantitative study of science, the Scientometrics (DE SOLLA PRICE, 1978); b) the "Research Evaluation" journal begins in 1991 (PAGE, 1991) and has become increasingly relevant with focus on societal emphasis about accountability and documenting the value of research.

Another mark date is 1994 when in "Scientometrics" journal devoted a special issue about the crisis of Scientometrics field that suffers one kind of fragmentation (GLÄNZEL; SCHOEPFLIN, 1994). They reflect on changes, on subjects, on stakeholders, and related background disciplines.

The emergence in the information market in 2004 of *Elsevier's Scopus* (with a multidisciplinary scope, greater coverage of scientific journals from different geographical regions and user-friendliness) is a milestone for the evaluation of research. The main advantages of this competitive factor of Scopus's entry become twofold: a) the sources of research from social sciences become more visible and b) the WoS platform was forced to take a continually improve its features and broaden the scope of its publication sources.

One limitation of this study is related to the database used. Although we used WoS in this case, it would be desirable for future efforts to complement data from diverse sources such as *Scopus* or *Scielo*, and incorporating new data every year to have an up-to-date work material.

Those insights allow inputs and motivation to go beyond this exploratory study. Our next step is to perform a deep integrative literature review using this study output as a starting input data. We will perform a co-word analysis of all keywords and a content analysis of the relevant and core scientific publication on Research Evaluation.

References

AKSNES, D. W. Characteristics of highly cited papers. **Research Evaluation,** Oxford, UK , v. 12, n. 3, p. 159-170, 2003.

AKSNES, D. W.; RIP, A. Researchers' perceptions of citations. **Research Policy,** Brighton, UK, v. 38, n. 6, p. 895-905, 2009.

ANDERSON, G. **Fundamentals of educational research**. London: Routledge, 1998.

APA. **Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)**. 4. Arlington: American Psychiatric Association, 1994.

BAIRD, L. M.; OPPENHEIM, C. Do citations matter? **Journal of Information Science,** Aberystwyth, UK, v. 20, n. 1, p. 2-15, 1994.

BARTOL, T. et al. Assessment of research fields in Scopus and Web of Science in the view of national research evaluation in Slovenia. **Scientometrics,** Belgium, v. 98, n. 2, p. 1491-1504, Feb 2014.

BECK, A. T. et al. An inventory for measuring depression. **Archives of General Psychiatry,** USA, v. 4, p. 561-571, 1961.

BORDONS, M.; FERNÁNDEZ, M.; GÓMEZ, I. Advantages and limitations in the use of impact factor measures for the assessment of research performance. **Scientometrics,** Belgium, v. 53, n. 2, p. 195-206, 2002.

BROUWERS, M. C. et al. AGREE II: advancing guideline development, reporting and evaluation in health care. **Canadian Medical Association Journal,** Canada, v. 182, n. 18, p. E839-E842, dec. 2010.

BUXTON, M.; HANNEY, S. How Can Payback from Health Services Research Be Assessed? **Journal of Health Services Research,** UK, v. 1, n. 1, p. 35-43, 1996.

CALLAHAM, M.; WEARS, R. L.; WEBER, E. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. **JAMA,** USA, v. 287, n. 21, p. 2847-2850, 2002.

CECI, S. J.; PETERS, D. P. Peer Review: A Study of Reliability. **Change: The Magazine of Higher Learning,** Canada, v. 14, n. 6, p. 44-48, set. 1982.

CHARNES, A.; COOPER, W. W.; RHODES, E. Measuring the efficiency of decision making units. **European Journal of Operational Research,** Europe, v. 2, n. 6, p. 429-444, 1978.

CLUZEAU, F. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. **Quality & Safety in Health Care,** UK, v. 12, n. 1, p. 18-23, 2003.

COATS, A. W. The role of scholarly journals in the history of economics: An essay. **Journal of Economic Literature,** USA, v. 9, n. 1, p. 29-44, 1971.

COHEN, J. **Statistical Power Analysis for the Behavioral Sciences**. 2. ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
CRONBACH, L. J.; MEEHL, P. E. Construct validity in psychological tests. **Psychological Bulletin,** USA, v. 52, n. 4, p. 281-302, 1955.

DE SOLLA PRICE, D. Editorial statements. **Scientometrics,** Belgium, v. 1, n. 1, p. 3-8, sep. 1978.

DIETZ, J. S.; BOZEMAN, B. Academic careers, patents, and productivity: industry experience as scientific and technical human capital. **Research Policy,** Brighton, UK,v. 34, n. 3, p. 349-367, 2005.

DIMAGGIO, P. J.; POWELL, W. W. The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. **American Sociological Review,** USA, v. 48, n. 2, p. 147-160, 1983.

DONOVAN, C. The qualitative future of research evaluation. **Science and Public Policy,** Oxford, UK, v. 34, n. 8, p. 585-597, 2007.

FREEMAN, L. C. Centrality in Social Networks Conceptual Clarification. **Social Networks,** UK, v. 1, n. 3, p. 215-239, 1979.

GARFIELD, E. Citation indexes for science: a new dimension in documentation through association of ideas. **Science,** UK, v. 122, p. 108-111, 1955.

GARFIELD, E. Citation Analysis As a Tool in Journal Evaluation - Journals Can be Ranked by Frequency and Impact of Citations for Science Policy Studies. **Science,** UK, v. 178, p. 471-479, 1972.

GARFIELD, E. Is citation analysis a legitimate evaluation tool? **Scientometrics,** Belgium, v. 1, n. 4, p. 359-375, May 1979.

GARFIELD, E. Current Contents. Mapping cholera research and the impact of Sambhu Nath De of Calcutta. **Essays of an Information Scientist**, USA, v. 9, n. 15, p. 112, 1986.

GARFIELD, E. **Citation indexes for retrieval and research evaluation**. Consensus Conference on the Theory and Practice of Research Assessment. Capri, Italy 1996.

GARFIELD, E. The evolution of the science citation index. **International microbiology,** UK, v. 10, n. 1, p. 65, 2007.

GARNER, J. G. *et al.* Assessing research network and disciplinary engagement changes induced by an NSF program. **Research Evaluation,** Oxford, UK, v. 21, n. 2, p. 89-104, June 2012.

GEARY, J.; MARRIOTT, L.; ROWLINSON, M. Journal Rankings in Business and Management and the 2001. Research Assessment Exercise in the UK. **British Journal of Management,** UK, v. 15, n. 2, p. 95-141, 2004.

GIBBONS, M. *et al.* **The New Production of Knowledge:** The dynamics of science and research in contemporary societies. London: SAGE Publications, 1994.
GLÄNZEL, W.; SCHOEPFLIN, U. Little scientometrics, big scientometrics ... and beyond? **Scientometrics,** Belgium, v. 30, n. 2, p. 375-384, 1994.

GLANZEL, W.; THIJS, B.; DEBACKERE, K. The application of citation-based performance classes to the disciplinary and multidisciplinary assessment in national comparison and institutional research assessment. **Scientometrics,** Belgium, v. 101, n. 2, p. 939-952, nov. 2014.

GLASER, B. G.; STRAUSS, A. L. **The Discovery of Grounded Theory:** Strategies for Qualitative Research. Chicago: Aldine, 1967.

GUTHRIE, S. *et al*. **Measuring research:** a guide to research evaluation frameworks and tools. Santa Monica: Rand, 2013.

HICKS, D. Evolving regimes of multi-university research evaluation. **Higher Education,** Oxford,UK, v. 57, n. 4, p. 393-404, Apr. 2009.

HICKS, D. Performance-based university research funding systems. **Research Policy,** Brighton, UK, v. 41, n. 2, p. 251-261, 2012.

HICKS, D. What year? Difficulties in identifying the effect of policy on university output. **Journal of Informetrics**, Belgium, v. 11, n. 3, p. 933-936, 2017.

HIRSCH, J. An index to quantify an individual's scientific research output. **Proceedings of the National Academy of Sciences of the United States of America,** USA, v. 102, n. 46, p. 16569–16572, 2005.

HOFSTEDE, G. **Culture's consequences:** International differences in work-related value. Beverly Hills, CA: Sage, 1980.

HOFSTEDE, G. **Culture's consequences: comparing values, behaviors, institutions and organizations across nations**. 2. ed. Thousand Oaks: SAGE Publications, 2001.

HOUSE, R. J. *et al*. **Culture, leadership, and organizations:** the GLOBE study of 62 societies. Thousand Oaks, CA: Sage, 2004.

JUBB, M. Prometheus assessed? Research measurement, peer review and citation analysis. **Learned Publishing,** Cmbridge,UK, v. 26, n. 3, p. 228-229, 2013.

KUHN, T. S. **The structure of scientific revolutions**. Chicago, US: University of Chicago Press, 1962.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, Washington, USA, v. 33, n. 1, p. 159-174, 1977.

LEE, S.; BOZEMAN, B. The impact of research collaboration on scientific productivity. **Social Studies of Science,** Canada, v. 35, n. 5, p. 673-702, 2005.

LEVIN, S. G.; STEPHAN, P. E. Research productivity over the life cycle: evidence for academic scientists. **The American economic review**, Nashville, USA, p. 114-132, 1991.

LEYDESDORFF, L. Theories of citation? **Scientometrics,** Belgium, v. 43, n. 1, p. 5-25, 1998.

LEYDESDORFF, L.; MILOJEVIĆ, S. Scientometrics *In*: INTERNATIONAL Encyclopedia of the Social & Behavioral Sciences (Second Edition). Oxford: Elsevier, 2015. p. 322-327.

LIEBOWITZ, S. J.; PALMER, J. P. Assessing the relative impacts of economics journals. **Journal of Economic Literature,** USA, v. 22, n. 1, p. 77-88, 1984.

LINCOLN, Y. S.; GUBA, E. G. **Naturalistic Inquiry**. Beverly Hills: SAGE Publications, 1985.

LOTKA, A. J. The frequency distribution of scientific productivity. **Journal of the Washington academy of sciences,** USA,v. 16, n. 12, p. 317-323, 1926.

MACDONALD, S.; KAM, J. Ring a ring o' roses: quality journals and gamesmanship in management studies. **Journal of Management Studies,** UK, v. 44, n. 4, p. 640-655, June 2007.

MACROBERTS, M. H.; MACROBERTS, B. R. Problems of citation analysis: A critical review. **Journal of the American Society for information Science,** USA, v. 40, n. 5, p. 342, 1989.

MAHONEY, M. J. Publication prejudices: An experimental study of confirmatory bias in the peer review system. **Cognitive Therapy and Research,** Boston, USA, v. 1, n. 2, p. 161-175, 1977.

MARTIN, B. R. The Research Excellence Framework and the 'impact agenda': are we creating a Frankenstein monster? **Research Evaluation,** Oxford, UK, v. 20, n. 3, p. 247-254, Sept. 2011.

MEHO, L. I.; YANG, K. 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,** USA, v. 58, n. 13, p. 2105-2125, 2007.

MERTON, R. **On theoretical sociology. Five essays, old and new**. New York: The Free Press, 1967.

MILES, M. B.; HUBERMAN, A. M.; SALDAÑA, J. **Qualitative Data Analysis**. Thousand Oaks: SAGE Publications, 2013.

MOED, H. F. **Citation analysis in research evaluation**. Dordrecht: Springer Verlag, 2005.

NALIMOV, V.; MULCJENKO, B. **Measurement of science. Study of the development of science as an information process**. Washington , DC: Foreign Technology Division, 1971.

NEDERHOF, A. J. et al. Research performance indicators for university departments: A study of an agricultural university. **Scientometrics,** Belgium,v. 27, n. 2, p. 157-178, 1993.

NEWMAN, M. E. The structure of scientific collaboration networks. **Proceedings of the national academy of sciences,** USA, v. 98, n. 2, p. 404-409, 2001.

NOWOTNY, H.; SCOTT, P.; GIBBONS, M. **Re-thinking science:** knowledge and the public in an age of uncertainty. Cambridge: Polity Press, 2001.

OECD. **The measurement of scientific and technical activities:** proposed standard practice for surveys of research and development. Paris: OECD, 1962.

OECD. **Frascati manual**. 6. ed. Paris, France: OECD, 2002.

OSWALD, A. J. An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-Makers. **Economica,** USA,v. 74, n. 293, p. 21-31, 2007.

PAGE, W. Research evaluation: welcome. **Research Evaluation,** Oxford, UK, v. 1, n. 1, p. 2-2, 1991.

PATSOPOULOS, N. A.; ANALATOS, A. A.; IOANNIDIS, J. P. Relative citation impact of various study designs in the health sciences. **JAMA - Journal of the American Medical Association,** USA, v. 293, p. 2362-2366, 2005.

PINHO, I .; ROSA, M. J. Towards a new research units' evaluation model in an open science context- a proposal. **Research Day Proceedings**, Aveiro, Portugal, 2016.

POLANYI, M. The republic of science: Its political and economic theory. **Minerva,** Netherlands, v. 1, p. 54-73, 1962.

PORTER, A. L. *et al*. Measuring researcher interdisciplinarity. **Scientometrics,** Belgium, v. 72, n. 1, p. 117-147, 2007.

STARBUCK, W. H. How much better are the most-prestigious journals? The statistics of academic publication. **Organization Science,** USA, v. 16, n. 2, p. 180-200, 2005.

ŠUBELJ, L.; VAN ECK, N. J.; WALTMAN, L. Clustering scientific publications based on citation relations: a systematic comparison of different methods. **PLoS ONE,** California, USA, v. 11, n. 4, p. e0154404, 2016.

VAN ECK, N. J.; WALTMAN, L. CitNetExplorer: a new software tool for analyzing and visualizing citation networks. **Journal of Informetrics,** UK, v. 8, n. 4, p. 802-823, Oct. 2014.

VAN ECK, N. J.; WALTMAN, L. **VOSviewer manual**. Leiden: University of Leiden, 2015.

VAN RAAN, A. F. J. Fatal attraction: conceptual and methodological problems in the ranking of universities by bibliometric methods. **Scientometrics,** Belgium, v. 62, n. 1, p. 133-143, 2005.

VINKLER, P. Relations of relative scientometric indicators. **Scientometrics,** Belgium, v. 58, n. 3, p. 687-694, Nov. 2003.

VOOS, H.; DAGAEV, K. S. Are All Citations Equal? Or, Did We Op. Cit. Your Idem? **Journal of Academic Librarianship,** UK, v. 1, n. 6, p. 19-21, 1976.

WALTMAN, L.; VAN ECK, N. J. A new methodology for constructing a publication-level classification system of science. **Journal of the American Society for Information Science and Technology,** USA, v. 63, n. 12, p. 2378-2392, 2012.

WEISS, C. H. The many meanings of research utilization. **Public administration review,** USA, v. 39, n. 5, p. 426-431, 1979.

WEISS, C. **Evaluation:** methods for studying programs and policies. 2. ed. Englewood Cliffs: Prentice Hall, 1998.

WOOLF, S. H. et al. Clinical guidelines: potential benefits limitations and harms of clinical guidelines. **BMJ,** USA, v. 318, p. 527-530, 1999.

YOUNG, N. S.; IOANNIDIS, J. P. A.; AL-UBAYDLI, O. Why current publication practices may distort science. **Plos Medicine,** UK, v. 5, n. 10, p. e201, 2008.