U.Porto Journal of Engineering, 6:2 (2020) 98-118
ISSN 2183-6493
DOI: 10.24840/2183-6493_006.002_0009

Received: 28 November, 2019
Accepted: 20 December, 2019
Published: 27 November, 2020

Visualising the Collaboration Network of a European Marine Research Infrastructure: A Bibliometric and Social Network Analysis

Ana Cavadas
CIIMAR - Interdisciplinary Centre of Marine and Environmental Research, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos, s/n, 4450-208 Matosinhos, Portugal; Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal (ana.cavadas@fe.up.pt) ORCID 0000-0002-4749-8200

Abstract
Collaboration has been recognised as an effective mean to foster the performance of research organisations. Throughout the years, numerous initiatives and investments have been made towards building research capacity and promoting collaborative research.
With a focus on the effects of the implementation of a distributed Research Infrastructure on Marine Sciences – the European Marine Biological Resource Centre (EMBRC-ERIC), this study applies bibliometric approaches and social network theory to examine the structure, characteristics and trends in the collaboration network of the European Marine Sciences research community over the last 20 years.

Author Keywords. Social Network Analysis, Bibliometric, Collaboration Network, Marine Sciences, EMBRC-ERIC.

Type: Research Article
Open Access Peer Reviewed

1. Introduction

As the world moves further into the age of “big science”, the production of vast knowledge has increasingly led towards the establishment of Research Infrastructures.

Research infrastructures (RIs) are designated by the European Commission as facilities, resources and services used worldwide by the research communities to conduct research and foster innovation. RIs are enablers of high-quality research, providers of advanced services and data as well as prescribers of leading-edge technologies, being privileged places where research meets innovation (ESFRI 2018):

- Highly skilled scientists, engineers, technicians and managers, funding agencies, public authorities, policy decision-makers and industry;
- Scientific and technical multi- and cross-disciplinarity and a mix of an extensive range of interactions with economic and societal surrounding environments.

Regardless of the research field, RIs operate as knowledge hubs where partner research institutions work together towards a common goal: to excel in a given scientific area. Furthermore, in the biennial review of the key trends in science, technology and innovation policy, published in 2016, the Organisation for Economic Co-operation and Development (OECD) has argued that the establishment of Research Infrastructures is a crucial strategy to foster a new era of “big science” by promoting the access to cutting-edge equipment, large-scale facilities and world-class knowledge (OECD 2016).

Adams (2012) highlighted that the rise of collaborative research networks is essential to tackle grand societal challenges as in physics, environment and health. RIs are core players in these networks, as they promote the involvement of large international teams, supported by major facilities and rich data for efficient knowledge transfer (Adams 2012). Building on this, and following the recommendations from the European Strategy Forum on Research Infrastructures, particular attention has been given to the discussion on the role of collaboration to ensure the long-term sustainability of research infrastructures (ESFRI 2018).

As the sustainability of the RIs in the globalising knowledge economy relies on their ability to effectively collaborate and attract new users, **collaboration** and **multidisciplinarity** have been point out as critical drivers for the long-term sustainability of Research Infrastructures (Florio 2019).

The European Marine Biological Research Centre (EMBRC-ERIC) is a distributed pan-European Research Infrastructure designed to foster marine biology and ecology research in notorious marine institutes across Europe. EMBRC-ERIC provides a unique access point that brings together around 20 leading marine stations and research centres across 24 locations in 9 different countries: Belgium, Greece, Spain, France, Israel, Italy, Norway, Portugal and the United Kingdom, as represented in Figure 1.

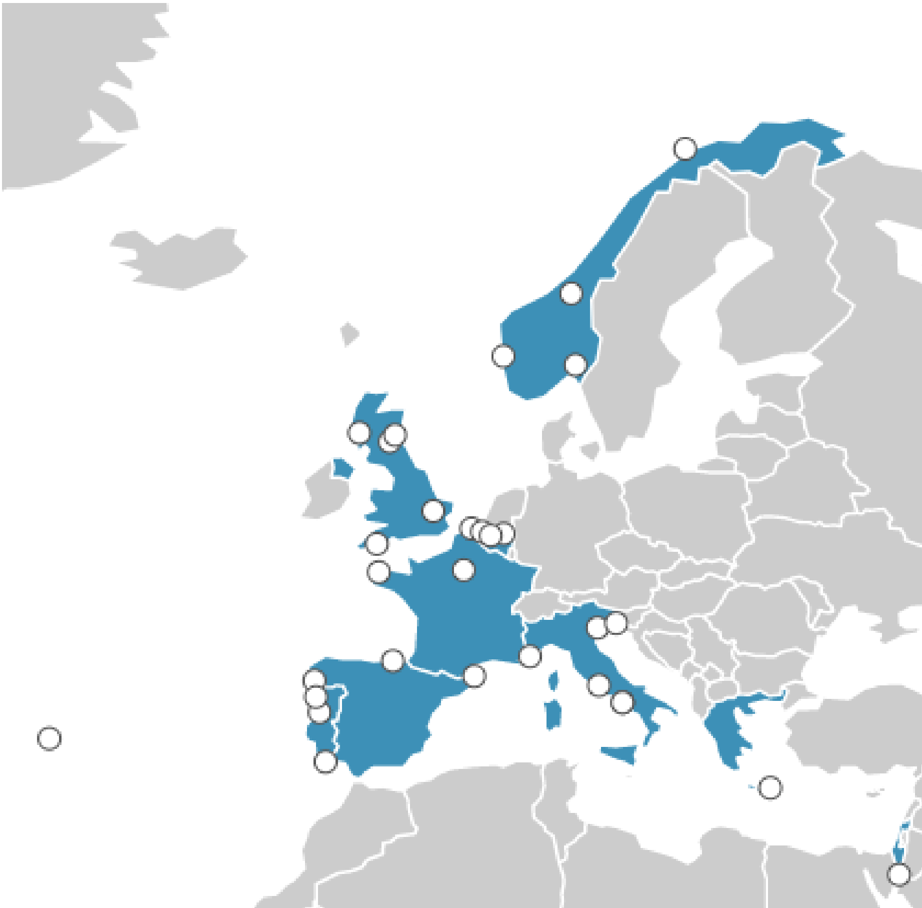

Figure 1: Map of the countries involved in EMBRC-ERIC. Circles represent the locations of the institutions involved. Map produced by the Flanders Marine Institute (VLIZ) and retrieved from EMBRC-ERIC webpage http://www.embrc.eu/partners

EMBRC-ERIC has become operational in 2018 and is projected to serve its users for the next 25 years. The long-term sustainability of EMBRC-ERIC is significantly connected to its capability to support the marine science community in fostering Blue Growth[1] – for that, EMBRC-ERIC must position itself as a global reference RI for marine sciences, by fostering networking and collaborative research (Nardello et al. 2017).

To tackle this strategic need, it is capital to explore and understand the structure of the collaboration network around EMBRC-ERIC, since it supports: recognition of leading researchers and organisations; mapping of prospective partners; research alignment and prioritisation (Liu et al. 2013); effectively designed and strategically planned and implemented collaborative research projects (Qiao, Mu, and Chen 2016); fostered cooperation and transnational research capacity building (Morel et al. 2009).

Chen, Zhang, and Fu (2019) have highlighted "International research collaboration" as an emerging research field within innovation studies, with an exponential increase in scientific publications over the last decade.

Vermeulen, Parker, and Penders (2013) reviewed the historical patterns of collaboration in life sciences. This well-structured study pin-pointed interesting shifts on the social dynamics of collaboration within the Biology research field throughout the years.

In the last decade, bibliometric and social network analysis have gained traction within research collaboration studies. Bender et al. (2015) applied Social Network Analysis to assess national and international collaborations of Germany-based researchers and research institutions working on neglected tropical diseases.

A study on the research performance of institutions in the field of nanoscience performed by Chinchilla-Rodríguez et al. (2018) highlighted the importance of examining the role of emerging countries in shaping the global collaboration network. More recently, Chinchilla-Rodríguez, Sugimoto, and Larivière (2019) analysed the correlation between nations leadership on collaboration networks with Research and Development investments.

Fonseca, Fernandes, and Fonseca (2017) examined the underlying network structure that describes the collaborative relations of Science and Technology organisations in the public sector. Furthermore, networks have been used to study the impact of collaboration on the performance of Korean public research institutions (Lee et al. 2012), to examine how collaboration networks influence the organisations' innovative performance (Guler and Nerkar 2012), to map the evolving theory, policy and practice over 40 years of coastal zone management research (Birch and Reyes 2018).

Building on this, this study aims to provide a comprehensive overview of collaboration patterns and major characteristics of the European collaboration network on Marine Sciences, in which all affiliated EMBR-ERIC institutions are included.

Bibliometrics and social network analysis (SNA) are applied to (i) comprehensively map the dynamics of knowledge production on Marine Sciences, (ii) further investigate the collaboration patterns, (iii) identify the key researchers and institutions, (iv) identify the core authors, influential journals, leading organisations/countries, (v) analyse keywords and co-citation clusters (vi) deduce emerging research topics, as well as (vii) examine the impact on scientific productivity of the institutions involved in a major distributed Research Infrastructure.

The findings of this study aim to contribute to the existing body of knowledge by highlighting the network dynamics within the European marine research community,

[1]Blue growth refers to a long term strategy recognised internationally to foster the sustainable growth and innovation potential of the marine and maritime sectors. More information is available on https://ec.europa.eu/maritimeaffairs/policy/blue_growth_en.

establishing its emerging research clusters, mapping the key institutions, and evidencing the benefit of international collaboration. Section 2 discusses the methodological background for this study, supporting the research approach applied, the search strategies, techniques and methods selected. Section 3 outlines the main findings and results of the bibliometric and social network analysis, discussing the leading research clusters; Section 4 outlines the conclusion and future directions.

2. Materials and Methods

2.1. Bibliometrics

Bibliometrics summarises a set of quantitative and statistical methods applied to analyse publication patterns, allowing to obtain insights from a macro perspective on the structure and collaborative relationships of scientific activities.

Bibliometrics has been widely used in information science, science and technology management, research performance evaluation, prediction and quantitative management. It has been applied mainly by universities and research centres, and also by policymakers, research directors, administrators and information specialists.

2.1.1.Descriptive bibliometrics

Paper Counts as a proxy to **Productivity.** Paper counts are a basic bibliometric measure providing the raw data for all citation analysis. Paper counts can be helpful to rank and compare the scientific production of different institutions. However, to effectively serve as a proxy to productivity, normalisation is required, and some additional data must be crossed, such as the number of researchers, funding availability, etc.

Average Citations per paper as a proxy to **Efficiency.** To weight the impact of research output, citations per paper is often used, being a useful metric whenever comparing large with small producers.

2.1.2.Bibliometric networks

Bibliographic Network Matrices containing the connected publication attributes are such as author(s), journal, keywords, publication date, etc. Rectangular matrices (Publications x Attributes) are used to represent connections of different attributes generate bipartite networks. Additionally, the references contained in scientific publications can generate a further network, based on co-citation or bibliographic coupling. These networks are analysed to capture meaningful properties of the underlying research system, and in particular to determine the influence of bibliometric units such as institutions, researchers and journals.

Co-authorship networks that can be classified based on (i) the units of the analysis i.e., what each node represents e.g., individuals, research teams, research organisations, countries; (ii) the type of information used to establish connections between nodes e.g., interactions or information shared; and (iii) the institutionalised domains to which the authors belong, focusing on the intra-organisational links.

Co-citation analysis enables the discovery of patterns of conceptual relationships among collections of co-cited documents. It forms clusters with citations that tend to be co-cited together frequently, and therefore are more likely to share related themes or concepts.

Co-occurrence analysis is instrumental when analysing the structure of the published literature in a specific research field, considering the relationships between keywords.

Multidimensional Scaling (MDS), Principal Component Analysis (PCA) or Multiple Correspondence Analysis (MCA) can be applied as dimensionality reduction techniques.

Bibliometric mapping is a quantitative method applied to visualise multiple bibliometric metrics of scientific publications. The produced networks can then be further explored through social network analysis.

2.2. Social network analysis

In the era of interconnectivity, social ties and connections are crucial determinants for the success of organisations and individuals. In this context, social network analysis (SNA) has been presented as a powerful tool to analyse the social phenomena and social structure of a community based on their relationships, activities and dynamics.

SNA is being applied as a convenient heuristic to map connections and quantify interactions between peers (Borgatti and Foster 2003). SNA typically analyses the structure of a network focusing on how functional entities are connected and what is their role in a network. Furthermore, one of the key aspects of SNA is the possibility to identify the node with the greatest influence and further explore the development of the network.

2.2.1. Network statistics

The generated network can be analysed in terms of density, centrality and cohesion. The structural properties of a network can be described by:

- **Size** referring to the number of vertices in the network.
- **Diameter** referring to the longest geodesic distance, i.e., the shortest path between each pair nodes, in the network.
- **Density** is the in the network, which is calculated referring to the number of connections the network compared to the total possible number of connections, measuring the degree of interconnectedness. Strong coordination between groups is visualised in a high-density network (Bodin, Crona, and Ernstson 2006). The computational formula for density is given by Equation 1:

$$Density = \frac{L}{[N(N-1)]} \quad (\mathbf{1})$$

Where N denotes the number of nodes and L denotes the number of connections (ties) in the network. Values ranging from 0 to 1, 0 indicates no connections and 1 indicates that all organisations within the network are connected.

- **Degree distribution** referring to the cumulative distribution of vertex degrees.
- **Degree centralisation** referring to the normalised degree of the complete network.
- **Average path length** referring to the average shortest distance between two vertices in a network.

2.2.2. Vertex statistics

"Centrality" is a key measure when analysing social networks, once it indicates the power and position of an individual or organisation in a social network. A network centrality can be explored in three major measures: **degree, closeness, and betweenness** (Freeman 1978).

To identify the core vertices in a network and to determine the tendency of two vertices to be both connected to a third vertex, the following measures can be computed:

Degree centrality measures the communication activity of a node, denoting the number of nodes adjacent to that node. This measure evaluates based on the number of direct connections, the degree to which the network is influenced by one or more nodes (e.g. authors, institutions or countries) (Poulin, Boily, and Mâsse 2000).

It is, therefore, useful to identify highly connected nodes, popular nodes that can rapidly connect with the broader network. Higher scores indicate influence by few individuals or institutions, and lower scores indicate a similar number of connections between all individuals or institutions in the network.

Closeness centrality is defined as the extent of the interconnectivity of a node with all other nodes in the network. This measure computes the shortest paths between all nodes, assigning each node a score based on their 'closeness'. This allows investigating how fast the information flows from one node to the others, being useful for finding the best-positioned nodes to rapidly influence the entire network (Newman 2005). Closeness centrality of a node is computed by Equation 2:

$$c_{Cl}(v) = \sum_{u \in V} \frac{1}{dist\ (v, u)} \tag{2}$$

Here, u is the focal vertex, v is another vertex in the network, and $dist\ (v, u)$ is defined as the shortest distance between the vertices u, v ∈ V.

Although closeness centrality could be useful to identify the major 'disseminators', in a highly connected network it is often found that all nodes have a similar score. Therefore, closeness will have greater utility when finding the influencers within a single cluster.

Betweenness centrality expresses the degree to which a few nodes, acting as "bridges", i.e., control the relationships of other nodes in the network. It can be defined as the ratio between the total shortest paths going through that node and all the possible shortest paths in the network and is computed by Equation 3:

$$c_B(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma\ (s, t|v)}{\sigma\ (s, t)} \tag{3}$$

Here, $\sigma\ (s, t|v)$ denotes the total number of shortest paths between s and t that pass through v, and $\sigma\ (s, t)$ defines the total number of shortest paths passing through v.

Betweenness is very useful for analysing collaboration dynamics, as higher scores indicate greater centralisation, suggesting that a small number of nodes hold authority over the network, or are controlling the collaboration ties between different clusters.

EigenCentrality measures the influence of a node considering how much a node is connected, i.e., how many ties it has to other nodes within the network. EigenCentrality takes degree centrality further by accounting for how well connected a node is, and how well connected are their connections, and doing this analysis throughout the entire network. Bonacich (1972) defined EigenCentrality measure as Equation 4:

$$c_{E_i}(v) = \alpha \sum_{\{u,v\} \in E} c_{Ei}(u) \tag{4}$$

Here, the vector $c_{E_i} = c_{E_i}(1), \ldots, c_{E_i}(N_v)^T$ is the solution to the eigenvalue problem $Ac_{E_i} = \alpha^{-1}\ c_{E_i}$, where A is the adjacency matrix for the network. According to Bonacich (1972) an optimal choice of α^{-1} is the largest eigenvalue of A.

Therefore, EigenCentrality may allow the identification of nodes with greater influence over the entire network by calculating the extended connections of a node.

PageRank is a variant of EigenCentrality, where a score is assigned to nodes based on their connections, and their connections' connections, it also considers the tie direction and weight – in the sense that influence is passed through the tie in only one direction, and each tie can give different amounts of influence.

Authority Score is a centrality measure where a high authority vertex is a vertex that is linked by several other vertices that are connecting several other vertices.

2.3. Research design

This bibliometric study includes descriptive statistics and uses text analytics and network analyses to reveal hidden patterns in the scientific publications from the Marine Sciences research community. The methodology presented in Figure 2 was designed as an end-to-end process, beginning with the dataset retrieval and ending with obtaining the findings from various data analysis and visualisation tools.

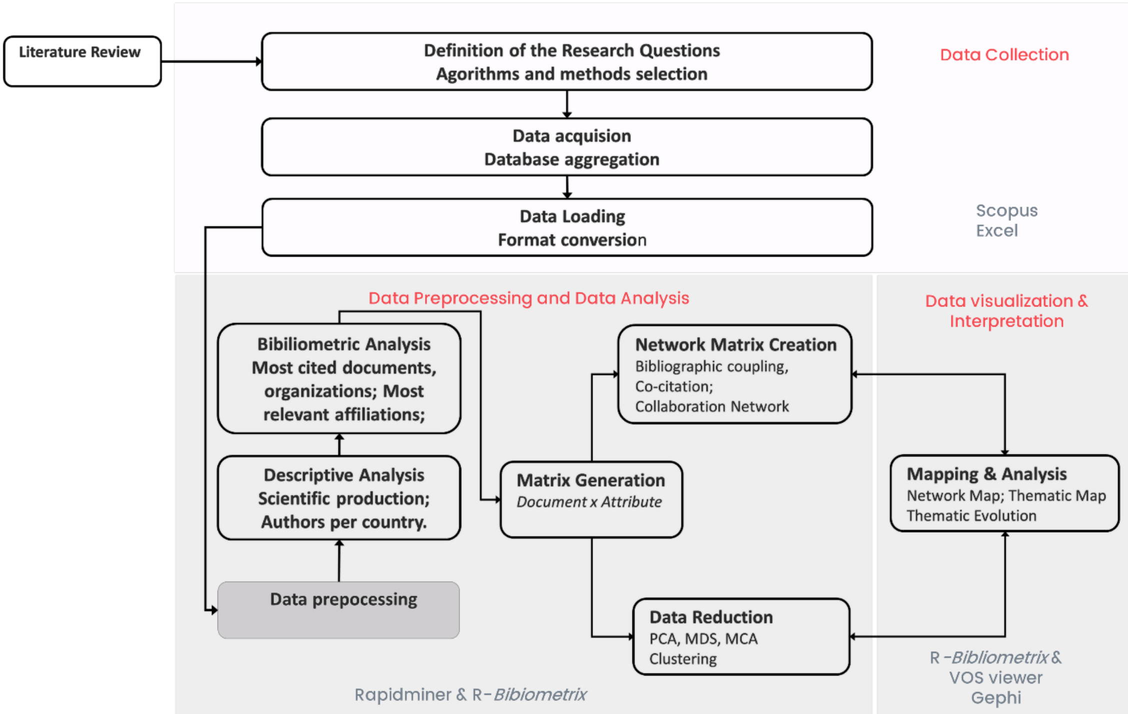

Figure 2: Methodological Research Framework for this study - workflow representation

2.4. Research questions

To select the appropriate data analysis tools and methods, it is important to have clearly defined research questions. Therefore, building on the general research question: *"How can Research Infrastructures extract greater knowledge from the interactions and relationships of research, in order to obtain meaningful insights on intellectual structure and collaboration behind a specific research field?",* and considering the scope of this study, three research questions were formulated and are stated in Table 1.

In this study, the Marine Science research community will be studied to get further information on the global scientific collaboration panorama around EMBRC-ERIC.

	Research Question	Method	Software	Knowledge contribution
(RQ1)	**What are the core organisations, countries and journals?**	Bibliometrics	RapidMiner, RStudio	Characterise the knowledge base of a Marine Sciences Research
(RQ2)	**What are the conceptual and intellectual structures within the European Marine Sciences scientific community?**	Bibliometrics and Social Network Analysis	RStudio, VOSViewer	Examine the conceptual and intellectual structure of European Marine Sciences research: co-occurrence network; bibliographic coupling
(RQ3)	**Which countries and institutions collaborate the most within the European Marine Sciences scientific community?**	Social Network Analysis	RStudio, VOSViewer, Gephi	Characterise the social structure behind the European Marine Sciences scientific community: Collaboration Network. Is there any evidence of the impact of the EMBRC-ERIC?

Table 1: Research questions, selected methods and software and the expected knowledge contribution from this study

2.5. Data collection

2.5.1.Source of data

For this study, scientific publication data was retrieved from Scopus Database (Elsevier BV Company, USA)[2] - the largest citation database of peer-review literature, providing further information such as the number of citations per each document, affiliation country for all the authors which will be of utmost relevance for this study.

2.5.2.Search strategy

The methodology was designed as an end-to-end process, from the dataset retrieval and to the visualisation and interpretation of the social network.

The Scopus database was used for the dataset retrieval, involving two steps: (i) an electronic database search for the most relevant text terms; and (ii) retrieving all records based on the identified text terms.

In step (i), the most frequent terms and its variants related to Marine Science research were identified from a sample of 2000 highly cited scientific papers containing "Marine" on its abstract, title or keywords. The title, abstract and keywords were retrieved and further processed with RapidMiner[3] (Mierswa et al. 2006) to obtain the scores Term Frequency – Inverse Document Frequency (TF-IDF). This text processing technique allowed the identification of the 10 most frequently appearing truncated terms. Thus, the following search query was defined and applied to obtain the dataset representative of the Marine Sciences research:

TITLE-ABS-KEY ("marine bio" OR "marine resourc*" OR "marine sci*" OR "marine product*" OR "marine pharma*" OR "marine drug" OR "marine env*" OR "marine eco*" OR "blue bio*" OR "blue eco*")*

[2]Scopus Bibliographic Database (Elsevier BV Company, USA) is available at https://www.scopus.com/.

[3]*RapidMiner Studio* 9.1 – Data science, machine learning, predictive analytics. Retrieved from https://rapidminer.com/.

For this study, only publications published in the English Language, with the involvement of European institutions over the last 20 years - from 1 January 1998 to 31 December 2018[4], were considered. Therefore, this additional query was applied:

*(LIMIT-TO (AFFILCOUNTRY, (*List of all European Countries*))) OR LIMIT-TO (PUBYEAR , 1998:2018)) AND (LIMIT-TO (LANGUAGE , "English"))*

The final set of records retrieved contained a total of 37839 publications.

2.6. Data preparation and pre-processing

All information collected was normalised and disambiguated, e.g. researchers' names, institution denominations, and any other information that appeared ambiguous. "Year" is an integer value, all the other fields containing text such as Title, Abstract and Keywords required pre-processing to generate standardised terms needed for subsequent analysis and visualisations.

2.7. Analytical tools

The methodological approach applied in this study is based on Bibliometrics. RapidMiner (Text Ming, data pre-processing), *Biliometrix* R-package (Bibliometric analysis), VOSViewer (Data visualisation) and Gephi (Data visualisation and Network statistics) were the software applied to analyse the publication data retrieved from the Scopus database.

The open-source *Bibliometrix* R-package[5] was used to perform the analysis coded according to the framework established by Aria and Cuccurullo (2017).

All generated networks were extracted in Pajek (*.net) format. The Pajek (*.net) files were then opened on VOSViewer (van Eck and Waltman 2014) and Gephi.

VOSviewer is a software tool developed by van Eck and Waltman (Leiden University) for constructing and visualising bibliometric networks based on Visualisation of Similarities (VOS) technology. VOSviewer allows through natural language processing techniques the creation of term co-occurrence maps based on textual data. The algorithms used in the software make the distinction between relevant and non-relevant terms, allowing the construction of relevant co-occurrence networks based on terms extracted from the dataset. The VOS technology enables the node of a dense network to be clearly displayed through interactions, being particularly interesting to map knowledge domains, especially when analysing large-scale data and constructing complex networks (van Eck and Waltman 2014).

Gephi – an open-source Java software package for network visualisation and analysis – was used for further visualisation whenever a more detailed analysis was required (Bastian, Heymann, and Jacomy 2009). Node centrality measures, including degree, density, betweenness, closeness, and eigenvector centrality, were measured with Gephi for each node and for all the network. ForceAtlas2 continuous graph layout algorithm was used to enhance network visualisation (Bastian, Heymann, and Jacomy 2009). By applying these software and inherent techniques structural proximities considering the connections between institutions were plotted into visual maps, allowing for a

[4]This time-frame was selected taking into consideration the main milestones regarding the establishment of EMBRC-ERIC: In 2008 EMBRC was first indicated in the Roadmap for European Research Infrastructures (ESFRI), being established as an operational Landmark in 2018. Therefore, in these 20 years we can analyse the collaboration prior to EMBRC-ERIC and the evolution during its preparation and implementation phase (2008-2018).

[5]More information on the *Biliometrix* R-package is available at http://www.bibliometrix.org/.

comprehensive interpretation of social dynamics within the Marine Sciences research community.

3. Discussion

The 37839 articles in the dataset involved 88376 authors, resulting in an average of 2.34 authors per article and 0.428 articles per author.

The descriptive analysis of the dataset indicates that collaboration is key among Marine Sciences research community, with only 3737 (10%) of the articles being published by a single author. Furthermore, other different co-authorship indices are shown in Table 2.

Period	**1998-2018**	**1998-2003**	**2004-2008**	**2009-2013**	**2014-2018**
Documents[(1)]	37839	5976	7312	10712	13839
Sources	4682	1110	1628	2085	2415
Keywords Scopus	91178	20139	33939	43724	44824
Author's Keywords	57976	11019	14454	22023	29282
Average citations per article	30.96	49.56	49.53	31.8	12.47
Authors	88376	15420	21408	33131	46425
Author Appearances	182406	21124	30768	52624	77890
Authors of single-authored documents	3017	886	803	807	824
Authors of multi-authored documents	85359	14534	20605	32324	45601
Single-authored documents	3737	1016	876	923	922
Documents per Author	0.428	0.388	0.342	0.323	0.298
Authors per Article	2.34	2.58	2.93	3.09	3.35
Co-Authors per Article[(2)]	4.82	3.53	4.21	4.91	5.63
Collaboration Index[(3)]	2.5	2.93	3.2	3.3	3.53

[(1)] All document types indexed in Scopus were considered. The distribution between document type is the following:

DOCUMENT TYPE	No.	%
ARTICLE	29321	77.49%
CONFERENCE PAPER	3918	10.35%
REVIEW	2344	6.19%
BOOK CHAPTER	1189	3.14%
EDITORIAL	281	0.74%
BOOK	204	0.54%
SHORT SURVEY	203	0.54%
NOTE	142	0.38%
LETTER	119	0.31%
ERRATUM	88	0.23%
ARTICLE IN PRESS	24	0.06%
ABSTRACT REPORT	3	0.01%
DATA PAPER	3	0.01%

[(2)] The Co-Authors per Articles index is calculated as the average number of co-authors per article.

[(3)] The Collaboration Index (CI) is calculated as Total Authors of Multi-Authored Articles/Total Multi-Authored Articles (Elango and Rajendran 2012; Koseoglu 2016)

Table 2: Descriptive analysis of the bibliographic data frame

3.1. Analysing the knowledge base of Marine Sciences (RQ1)

3.1.1. Evolution of the numbers of articles related to Marine Sciences

The number of scientific publications retrieved from the Scopus database related to Marine Sciences[6] for each year from 1998 to 2018 is presented in Figure 3. A total of 37839 documents were published over the last 20 years, revealing an increased interest and scientific production in the field of Marine Sciences. Figure 3 shows substantial growth in the number of publications between 2004 (1077 publications) and 2008 (1897 articles), this may be associated with the establishment under FP6 program of three Marine Networks of Excellence, namely EUR-OCEANS, MarBEF, Marine Genomics Europe. These networks were the starting point of ASSEMBLE in 2009 an FP6 Integrated Infrastructure Initiative project that created a network of leading European coastal marine biological research centres. The incentive from the European Commission in the scope of the EU Framework Programmes has supported the exponential rise of the open-access publications in recent years.

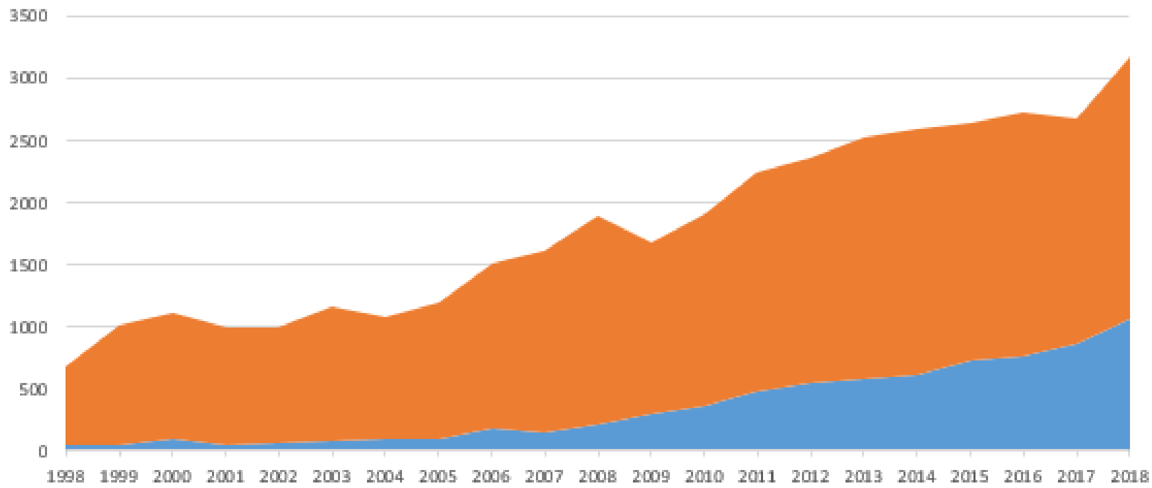

Figure 3: European Scientific Production
Number of Scopus publications within "Marine Sciences" field: 1998-2018

3.1.2. Identification of the leading countries and institutions contributing to the European Marine Sciences scientific production

To further understand the impact of the scientific productivity of European Marine Sciences research institutions, the lists of the most productive countries and the most productive institutions were retrieved and are presented in Figure 4 and Figure 5, respectively.

As shown in Figure 4, from 1998 to 2018, the most productive nation was the United Kingdom, being involved in 9822 publications, of which around 36% were co-authored with international peers. Germany is the second most productive country involved in 5944 publications of which 42% are in collaboration with international peers, followed by France with 5599 publications, 42% in collaboration with international peers.

With the exception of Israel, all the countries involved in EMBRC-ERIC are within the Top 15 most productive European countries.

[6]Publications intended to characterise the European scientific production within Marine Sciences were retrieved from Scopus applying search strategy stated in section 2.4.2.

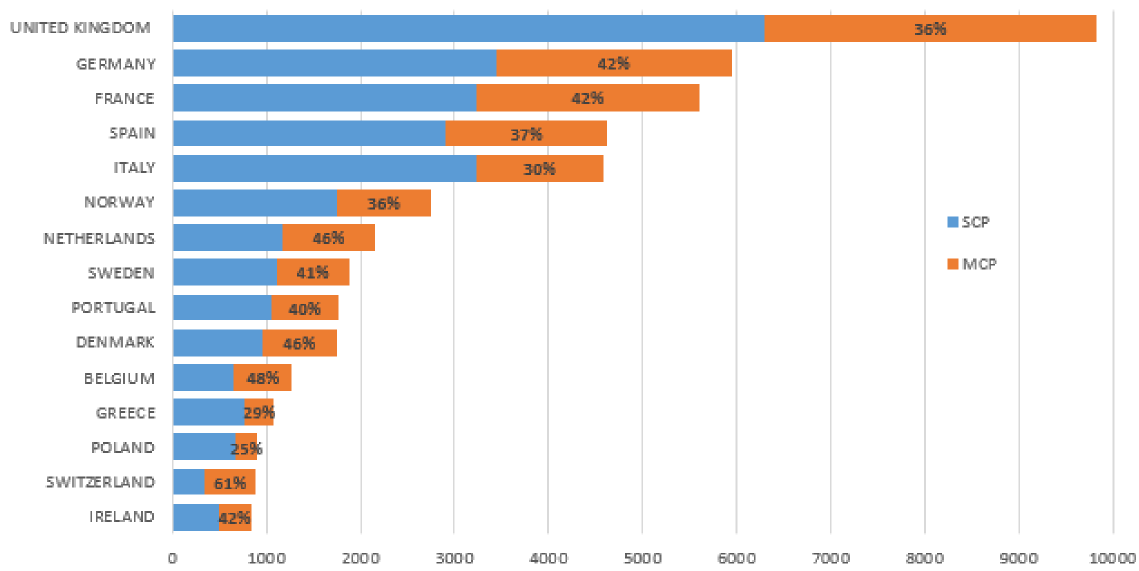

Figure 4: TOP 15 Most Productive Countries from 1998 to 2018
SCP: Single Country Publications; MCP: Multiple Country Publications
(representing the percentage of publications with international collaboration)

The most productive institution between 1998 and 2018 was the *Centre National pour la Recherce Scientifique (CNRS)* involved in 1842 scientific publications. CNRS is a research institution affiliated to EMBRC-France which constitutes the French component of EMBRC-ERIC. Figure 5 reports the publication dynamics of the top 30 institutions between 1998 and 2018, highlighting the institutions responsible for the main fluctuations in the scientific production in Marine Sciences.

3.1.3. Analysis of the most frequent publishing journals

The top five journals, publishing the majority of papers concerning to Marine Sciences were "JOURNAL OF THE MARINE BIOLOGICAL ASSOCIATION OF THE UNITED KINGDOM" (1140 articles, 3.9%), "MARINE POLLUTION BULLETIN" (1100, 3.8%), "HYDROBIOLOGIA" (854, 2.9%), "MARINE ECOLOGY PROGRESS SERIES" (694, 2.4%) and "PLOS ONE" (676, 2.3%). These five sources accounted for 15% of all peer-reviewed articles in the field of Marine Sciences. The 10 most frequently selected journals by the European Marine Sciences research community (Table 3) are the leading journals in the research domain, with average quartiles from the period between 1998 and 2008, distributed between Q1 and Q2. Furthermore, only two of the frequently selected journals are Open Access Journals, highlighting the potential for a shift towards more open research in the years to come.

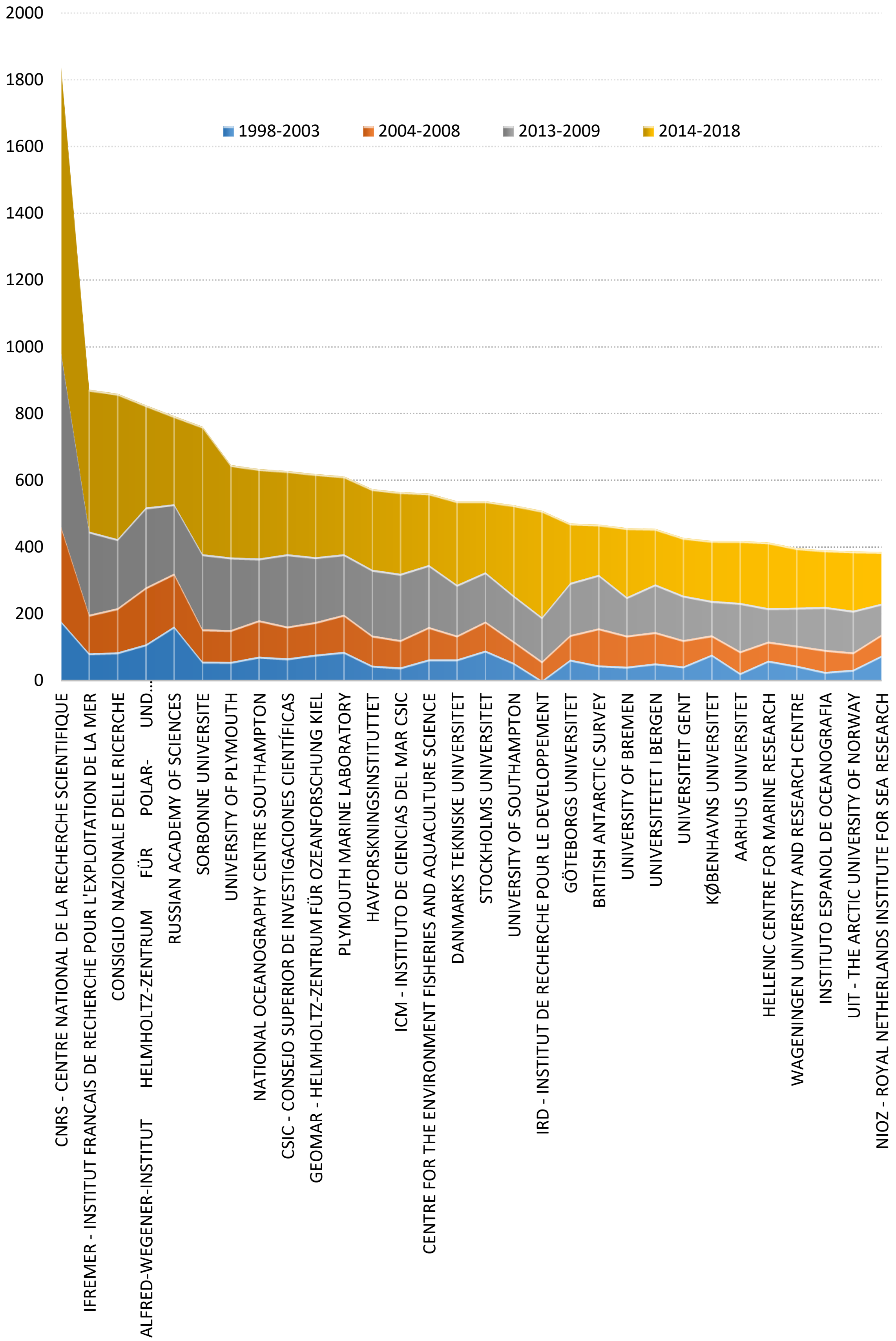

Figure 5: TOP 30 Most Productive Institutions from 1998-2018
Representation in 4 periods: 1998-2003, 2004-2008, 2013-2009, 2014-2018 to visualise changes in the publication dynamics

Sources (Journals)	No. Articles	Open Access Journal[1]	Quartile 1998-2018[2]
JOURNAL OF THE MARINE BIOLOGICAL ASSOCIATION OF THE UNITED KINGDOM	1140	NO	Q2
MARINE POLLUTION BULLETIN	1100	NO	Q1 (TOP10%)
HYDROBIOLOGIA	854	NO	Q2
MARINE ECOLOGY PROGRESS SERIES	694	NO	Q1
PLOS ONE	676	YES	Q1
MARINE ENVIRONMENTAL RESEARCH	509	NO	Q1 (TOP10%)
SCIENCE OF THE TOTAL ENVIRONMENT	448	NO	Q1 (TOP10%)
MARINE POLICY	441	YES	Q1 (TOP10%)
ICES JOURNAL OF MARINE SCIENCE	409	NO	Q1 (TOP10%)
JOURNAL OF MARINE SYSTEMS	381	NO	Q1

[1]Journal Open Access Policies retrieved from SHERPA/ROMEO – Publisher copyright policies & self-archiving: http://sherpa.ac.uk/romeo/index.php

[2]Average Quartile for the period 1998-2008 retrieved from Scimago Journal Ranking: https://www.scimagojr.com/

(TOP10%)Journal in the top 10% percentile of the highly-cited journals in the research field in 2018. Data obtained from CiteScore™ metrics powered by Scopus: https://www.scopus.com/sources

Table 3: TOP 10 Journals frequently selected for publications in Marine Sciences field

3.2. Analysing the conceptual and intellectual structures within Marine Sciences (RQ2)

3.2.1.Keyword co-occurrence

The Author's Keywords Tree Map was construed using *R-bibliometrix* software package. The 20 most frequent keywords extracted from the dataset and their respective countings are reported in Figure 6. Furthermore, *VOSviewer* was used to produce a keyword co-occurrence map (Figure 7). One of the great advantages of using the Natural Language Processing capabilities available on VOSviewer is related to the easiness of combining text mining, clustering and mapping algorithms, allowing to acquire more detail information on the relationships among the identified keywords.

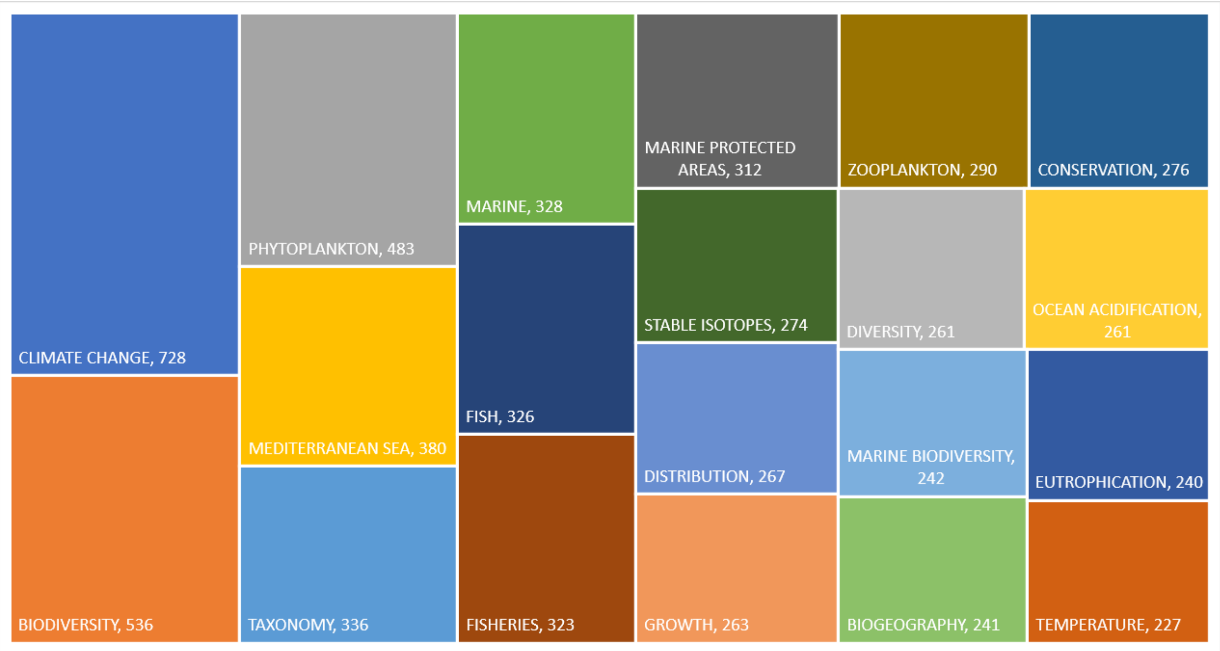

Figure 6: TOP 20 Author's Keywords Tree Map construed with R-bibliometrix

To obtain more information on these keywords, a density visualisation map was produced (Figure 7). The size of each point is dependent on the density of the number of items at that point. In the network density visualisation (Figure 7), keywords in the different clusters are displayed in different background colours. Keywords that are grouped into the same cluster, may reflect more related research areas. Therefore, from this analysis four main clusters within Marine Sciences research were identified: **1) Blue Biotechnology (in blue); 2) Marine Ecology (in green); 3) Marine Biodiversity (in red); 4) Marine Policy and Management (in yellow)**.

Climate Change has the more central positioning (interconnecting all four clusters) and the most frequent appearance (728 times). This indicates a transversal interest on the topic. Furthermore, this type of visualisation allows us to analyse the main topics within each cluster, for example, phytoplankton and eutrophication are the most frequent keywords related to the Marine Ecology cluster.

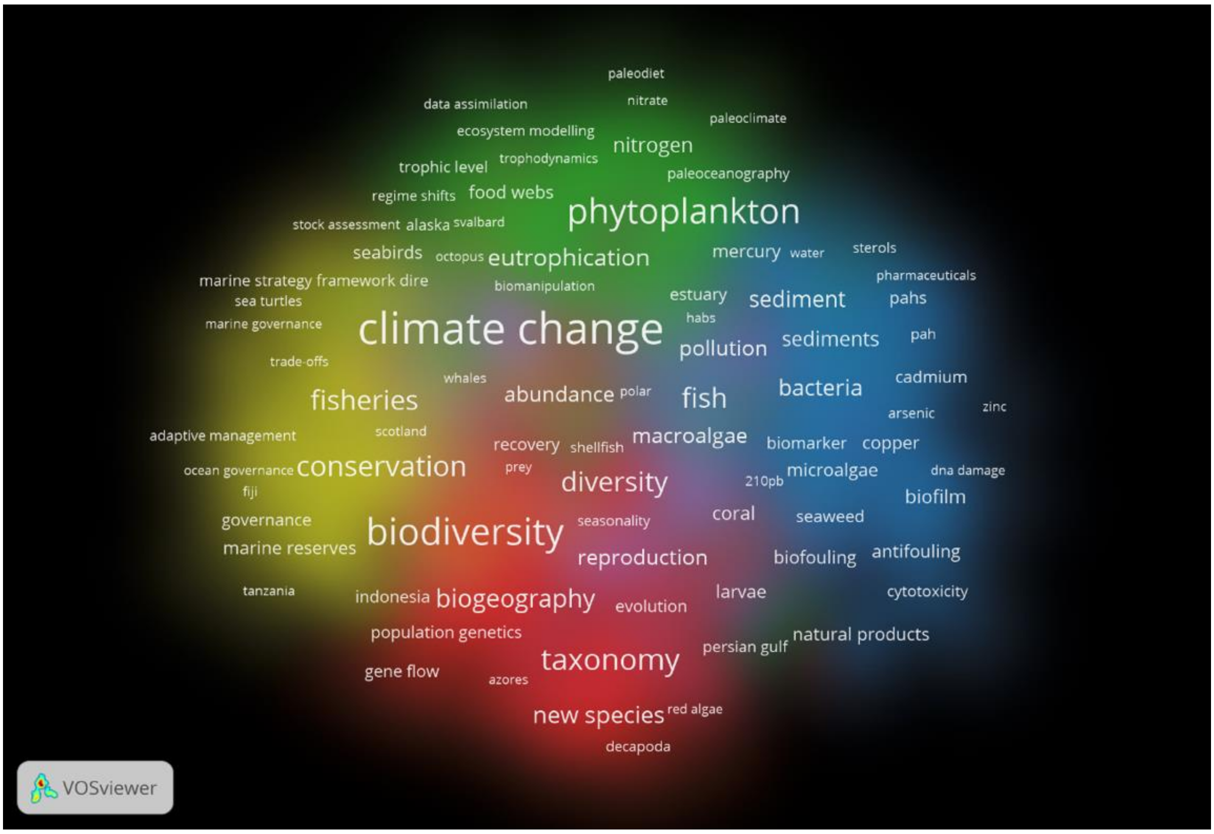

Figure 7: Authors' Keywords Network Density
Visualisation of four Clusters: Blue: Blue Biotechnology; Green: Marine Ecology;
Red: Marine Biodiversity; Yellow: Marine Policy and Management

3.3. Social network analysis of Marine Sciences research

3.3.1. Institution collaboration network visualisation and analysis (RQ3)

Institutional collaborations patterns have been examined with *Gephi* (Bastian, Heymann, and Jacomy 2009), to further understand how European institutions are evolving with the marine science research field.

To identify the core institutions main individual institutional centrality measures were computed: degree, Eigenvector, betweenness and closeness centrality. Degree centrality measures the frequency of connections that ties institutions within the network. In a weighted network, a high degree centrality indicates a highly connected institution, i.e., where high flows from and to an institution are visualised. Figure 8 presents the Collaboration Network of the institutions involved in scientific publications in the field of Marine Sciences from 1998 to 2018.

The collaboration network was visualised using Gephi. Nodes with betweenness centrality equal to zero were excluded (n=102). The network vertices (n=416) statistics were computed. Results are reported in Table 4. Collaborations in the scope of Marine Sciences are primarily affiliated to core institutions: (1) University of California (USA); (2) Institute of Marine Research (NOR); (3) Duke University (USA) and (4) Plymouth Marine Laboratory (GBR).

Centrality measures allow us to understand which institutions have more ties within a network and therefore, highlighting the most influential nodes within that network.

Furthermore, eigenvector along with betweenness centrality were measured to analyse if a node possesses the shortest path between other node pairs. Considering these measures, the ten main "*influencers*" are highlighted in Figure 8 - numbered from 1 to 10 (see Table 4).

Table 4 reports the vertices statistics of the 30 most relevant connectors within the network. It is shown that the University of California from the United States has the highest degree centrality with 119 network connections. It is followed by the Institute of Marine Research from Norway with 103 connections, and Plymouth Marine Laboratory from the United Kingdom with 96 connections. The analysis of the betweenness centrality showed the University of California and the Institute of Marine Research in Norway presented the highest scores, indicating their important role as "bridges", controlling the majority of relationships within the network.

A visualisation of clusters of four or more institutions was performed. The institutional clusters highlighted in Figure 8 showed a highly concentrated network, where the core institutions have interactions connecting the whole subnetwork.

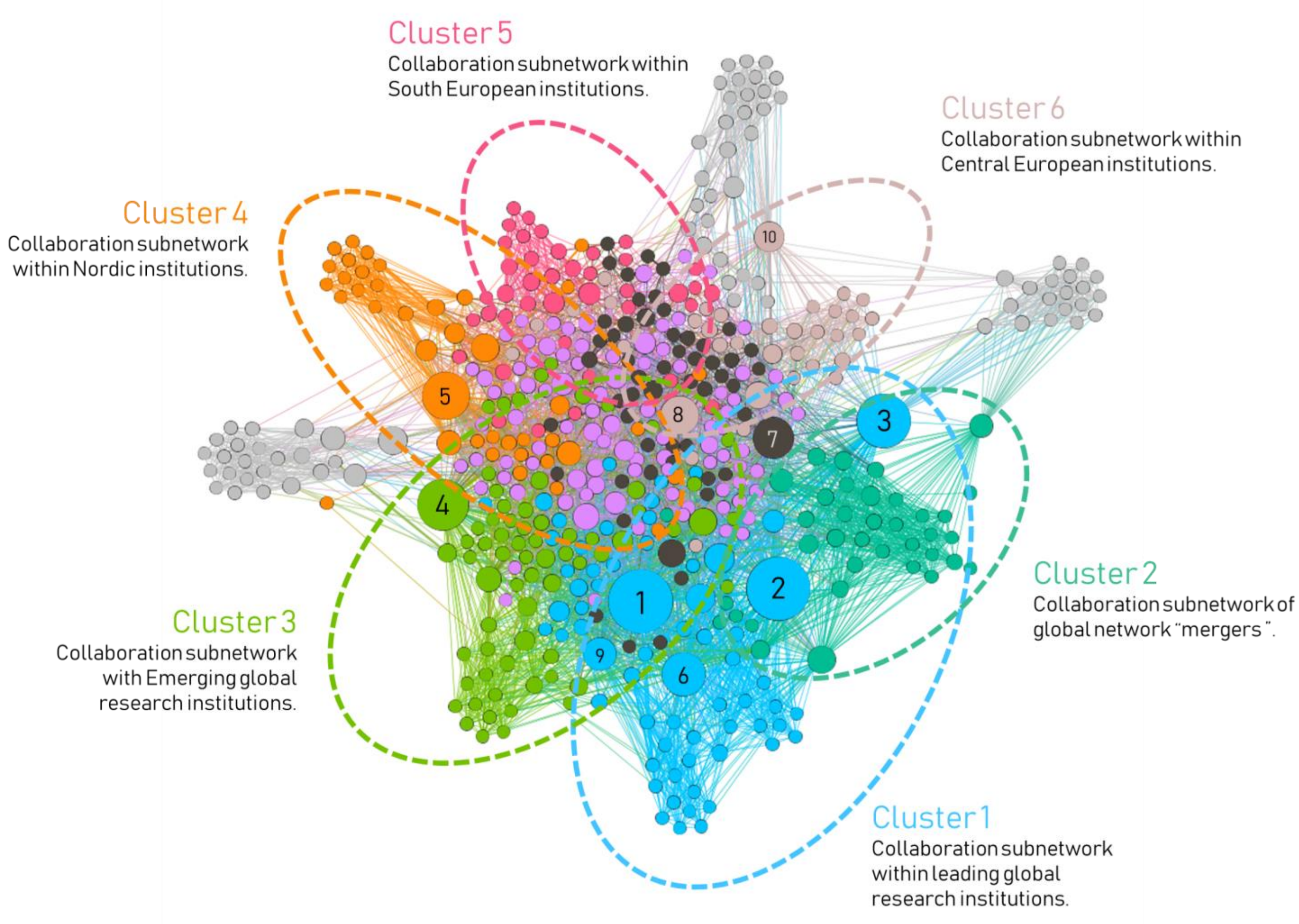

Figure 8: Institutions Collaboration Network within Marine Sciences field, 1998-2018
Network generated with R-bibliometrix, visualised and analysed with Gephi (Force Atlas 2 visualisation algorithm). Dataset with 502 nodes. Nodes with a betweenness score of zero were omitted. Node size represents betweenness centrality (larger nodes=higher scores). Node colour is representative of the cluster - Modularity class. Cluster 7 (brown) and 8 (pink) are central clusters (not highlighted). Stronger links are indicative of a higher frequency of collaboration between institutions. The core institutions are labelled from 1 to 10 (see Table 4), according to their betweenness centrality scores.

	Cluster	Institution	Degree Centrality	Weighted Degree	Closeness Centrality	Betweenness Centrality	EigenCentrality	Authority	Modularity class	PageRanks	Triangles
1	1	UNIVERSITY OF CALIFORNIA (USA)	119	171	0,557	8838	1,000	0,204	4	0,0098	996
2	1	INSTITUTE OF MARINE RESEARCH (NOR)	103	149	0,533	8695	0,861	0,175	4	0,0085	915
3	1	DUKE UNIVERSITY (USA)	91	113	0,498	7017	0,772	0,158	4	0,0071	860
4	3	PLYMOUTH MARINE LABORATORY (GBR)	96	121	0,526	6548	0,714	0,140	10	0,0081	651
5	4	UNIVERSITY OF OSLO (NOR)	84	106	0,501	5959	0,519	0,093	7	0,0075	525
6	1	UNIVERSITY OF WASHINGTON (USA)	95	133	0,528	5246	0,812	0,169	4	0,0077	828
7	7	UNIVERSITY OF SOUTHAMPTON (GBR)	78	101	0,513	4748	0,607	0,120	5	0,0069	430
8	6	UNIVERSITY OF BERGEN (NOR)	72	94	0,510	4393	0,549	0,107	1	0,0064	430
9	1	TECHNICAL UNIVERSITY OF DENMARK (DNK)	67	71	0,502	3444	0,529	0,107	4	0,0058	381
10	6	NATIONAL OCEANOGRAPHY CENTRE (GBR)	60	68	0,479	2936	0,456	0,086	1	0,0051	502
11	1	SCRIPPS INSTITUTION OF OCEANOGRAPHY (USA)	84	102	0,506	2835	0,783	0,164	4	0,0067	822
12	1	WOODS HOLE OCEANOGRAPHIC INSTITUTION (USA)	74	95	0,500	2796	0,665	0,140	4	0,0061	575
13	11	MACQUARIE UNIVERSITY (AUS)	37	39	0,439	2631	0,208	0,030	3	0,0032	262
14	4	NORWEGIAN POLAR INSTITUTE (NOR)	59	89	0,493	2567	0,401	0,071	7	0,0051	416
15	2	RUTGERS UNIVERSITY (USA)	72	75	0,487	2534	0,691	0,151	9	0,0055	743
16	7	STOCKHOLM UNIVERSITY (SWE)	58	74	0,491	2434	0,480	0,095	5	0,0052	267
17	3	UNIVERSITY OF HAWAII (USA)	66	72	0,491	2423	0,611	0,126	10	0,0054	600
18	6	UNIVERSITY OF OXFORD (GBR)	56	62	0,491	2333	0,442	0,086	1	0,0049	345
19	8	UNIVERSITY OF BRITISH COLUMBIA (CAN)	48	55	0,497	2245	0,393	0,077	0	0,0043	195
20	3	UNIVERSITY OF EXETER (GBR)	60	69	0,482	2069	0,490	0,099	10	0,0051	440
21	4	UNIVERSITY OF CAMBRIDGE (GBR)	52	55	0,470	1929	0,330	0,058	7	0,0045	311
22	4	AARHUS UNIVERSITY (DNK)	48	60	0,476	1912	0,349	0,067	7	0,0042	228
23	11	UIT - THE ARCTIC UNIVERSITY OF NORWAY (NOR)	32	35	0,420	1844	0,172	0,023	3	0,0028	248
24	2	UNIVERSITY OF ALASKA FAIRBANKS (USA)	55	58	0,475	1770	0,500	0,106	9	0,0043	552
25	2	UNIVERSITÉ DE MONTPELLIER (FRA)	51	51	0,424	1688	0,394	0,083	9	0,0039	599
26	11	UNIVERSITY OF TORONTO (CAN)	36	36	0,409	1615	0,162	0,021	3	0,0032	258
27	8	UNIVERSITY OF BRISTOL (GBR)	45	51	0,494	1567	0,374	0,071	0	0,0039	205
28	1	UNIVERSITY OF RHODE ISLAND (USA)	65	76	0,485	1536	0,622	0,130	4	0,0052	622
29	8	UNIVERSITY OF COPENHAGEN (DNK)	43	44	0,463	1521	0,274	0,052	0	0,0039	132
30	5	UNIVERSITAT DE BARCELONA (ESP)	35	42	0,422	1459	0,138	0,022	6	0,0035	135

Table 4: Institutional Collaboration Network – Network statistics calculated with Gephi

4. Conclusions

Researchers and research institutions have been incentivised to work in collaboration, and collaboration is being pointed out as the main driver for increased access to resources, skills, equipment, services, infrastructures, funding, leading to an enhanced research impact. When addressing complex multidisciplinary research topics, such as Marine Sciences, collaboration is an essential pillar for sustained knowledge creation.

Results indicate that to excel, Research Institutions within the network must actively work to establish new collaboration ties outside their immediate community. In Marine Sciences, and more specifically within EMBRC-ERIC context, higher performance can be achieved by increasing collaborations between researchers, academic and non-academic stakeholders.

The core institutions and nations as identified in this study are strategically positioned to lead future research efforts and to initiate interdisciplinary collaborations tackling the complex scientific challenges within Marine Sciences. In this context, EMBRC-ERIC may have an important role in structuring these networks on the regional, national and international level. When pursuing this challenge, there is a need for improved understanding of how the expansion of the research network may occur, how are collaborations formed, and how can the network structure be effectively managed towards maximised research performance and impact.

Furthermore, this study aimed to provide an overview of the current panorama of research collaboration in the field of Marine Sciences based on scientific publications. Nevertheless, it is important to note that scientific publication is not the only form of research collaboration. There are many other forms of scientific collaborations which do not result in co-authored peer-reviewed publications, such as shared supervision of PhD and MSc thesis, the preparation of research proposals, the involvement in joint R&D projects; patents; technology licensing; shared organisation of scientific conferences, etc. In this context, the findings from this study must be further examined in order to integrate more data into the analysis of the collaboration patterns between institutions and regions, namely by considering other types of data representative of these additional collaboration metrics.

References

Adams, J. 2012. "Collaborations: The rise of research networks". *Nature* 490, no. 7420 (october): 335-36. https://doi.org/10.1038/490335a.

Aria, M., and C. Cuccurullo. 2017. "bibliometrix: An R-tool for comprehensive science mapping analysis". *Journal of Informetrics* 11, no. 4 (november): 959-75. https://doi.org/10.1016/j.joi.2017.08.007.

Bastian, M., S. Heymann, and M. Jacomy. 2009. "Gephi: An open source software for exploring and manipulating networks". Paper presented in Proceedings of the Third International AAAI Conference on Weblogs and Social Media, San Jose, CA.

Bender, M. E., S. Edwards, P. von Philipsborn, F. Steinbeis, T. Keil, and P. Tinnemann. 2015. "Using co-authorship networks to map and analyse global neglected tropical disease research with an affiliation to Germany". *PLoS Neglected Tropical Diseases* 9, no. 12 (december): Article number e0004182. https://doi.org/10.1371/journal.pntd.0004182.

Birch, T., and E. Reyes. 2018. "Forty years of coastal zone management (1975–2014): Evolving theory, policy and practice as reflected in scientific research publications". *Ocean & Coastal Management* 153 (march): 1-11. https://doi.org/10.1016/j.ocecoaman.2017.12.003.

Bodin, Ö., B. Crona, and H. Ernstson. 2006. "Social networks in natural resource management: What is there to learn from a structural perspective?". *Ecology and Society* 11, no. 2 (december): Article number r2. https://doi.org/10.5751/ES-01808-1102r02.

Bonacich, P. 1972. "Factoring and weighting approaches to status scores and clique identification". *The Journal of Mathematical Sociology* 2, no. 1 (january): 113-20. https://doi.org/10.1080/0022250x.1972.9989806.

Borgatti, S. P., and P. C. Foster. 2003. "The network paradigm in organizational research: A review and typology". *Journal of Management* 29, no. 6 (december): 991-1013. https://doi.org/10.1016/s0149-2063(03)00087-4.

Chen, K., Y. Zhang, and X. Fu. 2019. "International research collaboration: An emerging domain of innovation studies?". *Research Policy* 48, no. 1 (february): 149-68. https://doi.org/10.1016/j.respol.2018.08.005.

Chinchilla-Rodríguez, Z., S. Miguel, A. Perianes-Rodríguez, and C. R. Sugimoto. 2018. "Dependencies and autonomy in research performance: examining nanoscience and nanotechnology in emerging countries". *Scientometrics* 115, no. 3 (june): 1485-504. https://doi.org/10.1007/s11192-018-2652-7.

Chinchilla-Rodríguez, Z., C. R. Sugimoto, and V. Lariviere. 2019. "Follow the leader: On the relationship between leadership and scholarly impact in international collaborations". *PLoS One* 14, no. 6 (june): e0218309. https://doi.org/10.1371/journal.pone.0218309.

Elango, B., and P. Rajendran. 2012. "Authorship trends and collaboration pattern in the marine sciences literature: A scientometric study". *International Journal of Information Dissemination and Technology* 2, no. 3: 166-69.

ESFRI - European Strategy Forum on Research Infrastructures Innovation Working Group. 2018. *Innovation-oriented cooperation of Research Infrastructures*. ESFRI Scripta Volume III. Dipartimento di Fisica - Università degli Studi di Milano.

Florio, M. 2019. *Investing in science: Social cost-benefit analysis of research infrastructures*. MIT Press.

Fonseca, B. P. F., E. Fernandes, and M. V. A. Fonseca. 2017. "Collaboration in science and technology organizations of the public sector: A network perspective". *Science and Public Policy* 44, no. 1 (february): 37-49. https://doi.org/10.1093/scipol/scw013.

Freeman, L. C. 1978. "Centrality in social networks conceptual clarification". *Social Networks* 1, no. 3: 215-39. https://doi.org/10.1016/0378-8733(78)90021-7.

Guler, I., and A. Nerkar. 2012. "The impact of global and local cohesion on innovation in the pharmaceutical industry". *Strategic Management Journal* 33, no. 5 (may): 535-49. https://doi.org/10.1002/smj.957.

Koseoglu, M. A. 2016. "Mapping the institutional collaboration network of strategic management research: 1980–2014". *Scientometrics* 109, no. 1 (october): 203-26. https://doi.org/10.1007/s11192-016-1894-5.

Lee, D. H., I. W. Seo, H. C. Choe, and H. D. Kim. 2012. "Collaboration network patterns and research performance: the case of Korean public research institutions". *Scientometrics* 91, no. 3 (june): 925-42. https://doi.org/10.1007/s11192-011-0602-8.

Liu, X., Z. Guo, Z. Lin, and J. Ma. 2013. "A local social network approach for research management". *Decision Support Systems* 56, no. 1 (december): 427-38. https://doi.org/10.1016/j.dss.2012.10.055.

Mierswa, I., M. Wurst, R. Klinkenberg, M. Scholz, and T. Euler. 2006. "YALE: Rapid prototyping for complex data mining tasks". In *KDD '06: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 935–40. https://doi.org/10.1145/1150402.1150531.

Morel, C. M., S. J. Serruya, G. O. Penna, and R. Guimarães. 2009. "Co-authorship network analysis: A powerful tool for strategic planning of research, development and capacity building programs on neglected diseases". *PLoS Neglected Tropical Diseases* 3, no. 8 (august): Article number e501. https://doi.org/10.1371/journal.pntd.0000501.

Nardello, I., A. Villanueva, B. Kloareg, W. H. C. F. Kooistra, M. Borra, L. Salvaudon, S. Berkowicz, and M. Merciecca. 2017. *EMBRC-ERIC Business Plan*. Paris, France: European Marine Biological Resource Centre. https://www.embrc.eu/newsroom/publications/embrc-business-plan-nardello-et-al-april-2017.

Newman, M. E. J. 2005. "A measure of betweenness centrality based on random walks". *Social Networks* 27, no. 1 (january): 39-54. https://doi.org/10.1016/j.socnet.2004.11.009.

OECD. 2016. *OECD Science, Technology and Innovation Outlook 2016*. Paris: OECD Publishing. https://doi.org/10.1787/sti_in_outlook-2016-en.

Poulin, R., M. C. Boily, and B. R. Mâsse. 2000. "Dynamical systems to define centrality in social networks". *Social Networks* 22, no. 3 (july): 187-220. https://doi.org/10.1016/s0378-8733(00)00020-4.

Qiao, L., R. Mu, and K. Chen. 2016. "Scientific effects of large research infrastructures in China". *Technological Forecasting and Social Change* 112 (november): 102-12. https://doi.org/10.1016/j.techfore.2016.07.029.

van Eck, N. J., and L. Waltman. 2014. "Visualizing bibliometric networks". In *Measuring scholarly impact*, edited by Y. Ding, R. Rousseau and D. Wolfram. Cham: Springer. https://doi.org/10.1007/978-3-319-10377-8_13.

Vermeulen, N., J. N. Parker, and B. Penders. 2013. "Understanding life together: A brief history of collaboration in biology". *Endeavour* 37, no. 3 (september): 162-71. https://doi.org/10.1016/j.endeavour.2013.03.001.