Scientometric analysis of the CHAOS journal (1991-2019): from descriptive statistics to complex networks viewpoints

Stéphanie Depickère,[1, a)] Jürgen Kurths,[2, 3, 4, b)] and Gonzalo Marcelo Ramírez-Ávila[1, c)]

[1)] *Instituto de Investigaciones Físicas, Universidad Mayor de San Andrés, Casilla 8635. La Paz, Bolivia*

[2)] *Potsdam Institute for Climate Impact Research (PIK), 14473 Potsdam, Germany*

[3)] *Department of Physics, Humboldt University, 12489 Berlin, Germany*

[4)] *Department of Control Theory, Nizhny Novgorod State University, 606950 Nizhny Novgorod, Russia*

We performed a scientometric analysis of CHAOS papers from 1991 to 2019, applying a careful disambiguation process for identifying the authors correctly. Firstly, we used standard scientometric tools based on descriptive statistics. This analysis enabled us to compute productivity and the degree of collaboration. The evolution in the number of authors, countries, and topics per article has an increasing trend. An analysis of the citations considering their temporal mean number exhibits a growing tendency in time. Secondly, we dealt with Lotka-Zipf's law considering the rank-distributions of 15 datasets. We found that the sum of Crossref citations by country was the only dataset for which the power-law was the only plausible distribution. Next, we examined the networks of authors, countries, and topics, going from the simplest case of undirected and unweighted networks to the general case of weighted and directed networks and assigning a weight to the individual nodes. Based on the networks' topology and features, we introduced diversity, collaboration, influence, and productivity measures and found a significant increase in the diversity of all the considered networks (authors, countries, and topics), but manifesting a very different network structure. The computation of typical network quantities combined with the communities' identification reveals the presence of several hubs and the existence of various communities that encompass nodes of all the continents in the case of countries. Finally, using the most general networks, it was possible to compute influence and productivity indexes to find the USA, China, and Germany's leadership inside the network.

PACS numbers: 89.75.-k, 02.10.Ox, 05.50.+q, 89.65.-s
Keywords: Complex systems; Graph theory; Ising model; Social systems; Bibliometrics.

[a]stephanie.depickere@gmail.com
[b]kurths@pik-potsdam.de
[c]gramirez@ulb.ac.be

In the last years, scientometrics turned into one essential tool to evaluate scientific production's impact, one approach to elaborate policies for improving scientific advances and distribute the funds efficiently to research institutions and projects that could potentially impact. Scientific publications are crucial to endorse work efficiency, the quality of research, and the collaborations among scientists, institutions, and countries. During the last two decades, complex network analysis emerged as a fundamentally important tool since almost all systems can be represented by complex networks regardless of their nature. In this work, we combine scientometric measures with complex network analysis from different standpoints, including the most general descriptions such as the concept of time-variable weighted networks and the nodes' relevance even if they might be isolated. The features mentioned above could be the basis for further advances in scientometrics and complex networks theory. The network analysis' main results indicate that diversity and collaboration have increased markedly, and the leadership in influence and productivity of the USA, China, and Germany within CHAOS.

I. INTRODUCTION

Nowadays, scientists must publish the results issuing from their research, best in distinguished peer-review journals. The publication record is then used as one basic criterion to evaluate scientists' qualities. Indeed, such an assessment is challenging and inevitably charged with subjectivity; it is why, some years ago, it was considered that these aspects were strictly qualitative. An important fact to measure the quality and scientific impact of countries, institutions, publications, and individuals has become to resort to scientometrics, which is defined as the study of the quantitative aspects of the scientific publishing as a communication system[1], where this system might be even involved in economic activity[2]. Scientometrics is also considered as part of the sociology of science, and it is widely used in the formulation of scientific policies. Scientometrics is called the science of sciences not in the sense that it is superior to other disciplines, but because it emphasizes on all related to science[3]. From a more technical viewpoint, scientometrics applies methods and tools of statistics to bibliometric databases to compute scientific and technological productivity indexes as well as the innovative degree that allow evaluating the impact of such works[4]. Considering these aspects, scientometrics might also determine the importance of collaborations, trends in scientific research, and the evolution of science and technology. As an attempt to quantify, it is possible to define basic categories and to introduce two basic units in evaluative scientometrics, namely, the research article and the citation as the units of information and impact respectively[5]. Recently, there has been a huge development in what concerns scientometrics; specific journals are devoted to this topic and even there are scientometric analyses about publications on scientometrics[6]. In the quantification process of scientific research and its impact, there are other fields related to scientometrics such as bibliometrics[7,8], informetrics[2], cybermetrics[1], and altmetrics[9–11].

Scientometrics deals with productivity and impact indicators. For the first case, Zipf's ideas[12] constitute a beacon for the analysis, mainly in what concerns the rank distributions. In the second case, it is essential to introduce citation patterns associated with their impact through quantities such as the h-index that represents the coincidence of the number of articles h having the same number of citations[13]. Evidently, the h-index does not represent fairly the impact of the authors, institutions, or countries, and some heeds should be considered in its use[14]. Some other attempts were proposed to improve the h-index[15,16]. Additionally, a new index g was introduced[17], defined as the highest rank in such a way that the g first publications have together, at least g^2 citations; consequently, g-index$\geq$ h-index. Later, it was developed an extension of the h and g-indexes to fractional and complementary indicators: a, m, and r, being respectively, the average citations of publications ($a = \langle \text{citations} \rangle$), the quotient between the h-index and the number of years since the first publication ($m = h/\text{yr.}$), and the square root of the sum of the citations of the publications ($r = \sqrt{\sum_{i=1}^{NP_h} Cit_i}$), where NP_h is the total number of papers comprised in the h-index kernel. Another, more refined, definition of the r-index considers the accumulated impact factor (IF) corresponding to the sum of the impact factors of the journals in which an article was cited and the number of citations (Cit) each article had[18] $r = \sum_{i=1}^{NP_h} IF_i \times Cit_i$, whose objective is to identify the potential quality of a work. The above indicators are well explained by different authors[19,20]. Thus, the combined h and r indexes give us a clearer picture of the quality of the work and the inclusion of m allows to evaluate the persistence of the quality throughout the time.

The study of scientometrics often used as a powerful tool for elaborating scientific policies and evaluating the scientific productivity of researchers, institutions, and countries is in expansion. Numerous studies about that have been prepared for the evaluation of specific topics such as in nanoscience, pharmacology, and statistics[21]. The evaluation of scientific production has also been studied for different countries[22,23] using the h-index on the journals covered by the Science Citation Index (SCI). An extensive comparison for 95 countries was done, where a paper's significance has been measured in terms of the number of citations it received during five years in SCI[24].

For all the aspects mentioned above, it is clear that scientometry is of capital importance. However, we have to emphasize the lack of completely objective indicators, sometimes there are situations that can lead to misleading results to the detriment of scientific progress[25].

We take here the journal CHAOS as a paradigmatic example to perform a scientometric analysis. This study is conducted using two perspectives. In the first one, we use the traditionnal scientometric tools, such as standard statistics for defining quantities related to collaboration and productivity. We also include a section dedicated to research the existence of power-laws in some rank-frequency distributions, following the Zipf's ideas that gave rise to the first scientometric study based on very simple statistical quantities. In the second perspective, we consider CHAOS as an evolving system in which the constituents and their interactions can give scientometric information, allowing us to study it from a complex network perspective. Indeed, we introduce normalized quantities based on the analysis of complex networks going from the simplest case (symmetric unweighted and undirected adjacency matrix) which allowed us to describe the evolution of diversity in CHAOS to the general case (asymmetric weighted and directed adjacency matrix, including the consideration of the nonzero diagonal) utilized to characterize the productivity. We show that measures based on networks contain more information in what concerns the scientometric variables because we identify clearly the nodes and edges of the networks. Thus, the relevant measures of collaboration, influence and productivity depend on the interaction of the nodes, featured by the links between them. This approach is a complement to the statistical methods that offer a global view of the above mentioned measures but do not allow to deepen in the structural aspects of the system. The present work considers as relevant parameters the type of article to which a certain weight is assigned depending on its feature (we bear in mind 14 types of articles). A weight is also assigned to authors, and countries according to their participation and relevance in each publication. Thus, the article is structured as follows: in Sect. II, we explain the method, including the data collection and their analysis; hence, a statistical analysis of all papers in CHAOS is performed considering common descriptive statistics for the parameters mentioned above that evolve in the course of the issues published in CHAOS. Subsequently, in Sect. III, rank-frequency distributions were calculated for 15 different datasets, including the number of publications by author, country, and topics; besides the sum of Crossref and ISI citations by author, country, and topics among others. After a thorough analysis, we determine the cases in which power-laws describe these distributions, verifying the relevance or not of the Zipf's law. An analysis based on complex networks is described in Sect. IV. In this analysis, different characteristics of the associated networks are quantified, considering such networks from four points of view: weighted and unweighted symmetric networks, as well as weighted in links and nodes asymmetric networks. These diverse network topologies allow us to define some concepts of scientometrics such as diversity, collaboration, influence, and productivity in terms of complex networks. Finally, in Sec. V, some conclusions and perspectives are presented.

II. DATA, METHOD AND STATISTICAL ASPECTS

In this Section, we describe the data collection, the methods allowing their analysis, and their most relevant aspects.

A. Data

CHAOS published four issues a year from 1991 to 2014, and from 2015 onwards, twelve issues a year. The raw data were obtained from the website of CHAOS by webscraping, using all the issues from Volume 1 to 29, corresponding to the whole period 1991–2019. The name of the authors and their respective affiliations were extracted, together with the year, the volume, and the issue numbers, as well as the topics associated with each article. The impact indexes (Crossref and ISI metrics), available on the CHAOS website, were also registered for each paper, together with the type of article. All the data were obtained using R on 19th July 2020.

According to the database, 14 types of articles were found: Announcement (16 items), Article-commentary (18 items), Fast track or Brief-report (65 items), Case-report (13 items), Correction (32 items), Editorial (10 items), Erratum (3 items), Introduction (49 items), Letter (1 item), Obituary (1 item), Other (68 items), Reply (4 items), Retraction (3 items), Research-article (4816 items), and Review-article (17 items). The items corresponding to Retraction and their original articles were removed from the databases. We considered that Correction and Erratum were just a complement to a published article, so these types and their corresponding items were also removed. Introduction corresponded to three different types of items: Introduction to a focus issue (41 items), Introduction to an image gallery (7 items), and Foreword (1 item), being the latter a short introduction to a special issue dedicated to an eminent scientist. The categories of Announcement, Foreword, and Obituary were not taken into account in the analysis, being a bit too far from a research article, and they were then removed from our databases. The category Other corresponded in fact to images and was then renamed as Image, except one element that was a Referee Acknowledgment and was deleted. Finally, in the cleaned databases, articles could belong to 11 types: Article-commentary (18 items), Fast track (65 items), Case-report (13 items), Editorial (10 items), Image (67 items), Introduction to images (7 items), Introduction to focus Issue (41 items), Letter (1 item), Reply (4 items), Research-article (4813 items), and Review-article (17 items). From the features of each type of article, we gave a weight representing their relative importance in terms of scientific results, issued from a systematic work including bibliographic research and formulation of conclusions (Table I). This weight was also used to calculate the relative importance of the authors and countries, distributing equally the article weight among these actors. This implies that the relative weight of an author (or a country) is related to the number of participating authors (countries) in a paper. We must point out that the weight of an author (or a country) is cumulative. Thus, if an author (country) x participated in a number of papers given by $np^{(x)}$ and with a weight for each paper given by $wp_i^{(x)}$, the resultant weight of an author (country) is then expressed as $w^{(x)} = \sum_{i=1}^{np^{(x)}} \frac{wp_i^{(x)}}{n_i^{(x)}}$, where $n_i^{(x)}$ is the number of authors (countries) participating in the publication i in which x takes part.

TABLE I. Type of articles, their assigned weight and their quantity during the whole period 1991-2019.

Type of article	Weight	Quantity
Commentary	0.10	18
Case-report	1.00	13
Editorial	0.25	10
Fast track	1.00	65
Image	0.10	67
Intro image	0.05	7
Intro focus issue	0.50	41
Letter	1.00	1
Reply	0.10	4
Research	1.00	4813
Review	1.00	17

A meticulous disambiguation process has been carried out for identifying correctly the authors.

Thus, the author database was cleared in order to keep only one writing name for the same author. Authors with similar names were looked for on the web via personal webpage, or page with an available publication list (e.g., ResearchGate, Google Scholar, etc) to confirm the presence of one or different authors. This process is essential to avoid mistaken results due to the insertion of errors into network data[26].

The database of the countries was built extracting the information from the affiliation or adding it in case of not being directly available. The country was attributed according to the official list of countries, with the ISO Alpha-2 code (https://www.nationsonline.org/oneworld/country_code_list.htm). The affiliation of the author of one Editorial item was lacking, the corresponding information was then added to the database. In case of multiple affiliations, we considered all of them when corresponding to different countries. If some affiliations of one author corresponded to the same country, we counted only once the country's participation in the network.

The raw list of topics includes topics, special topics, and collections. It was directly used in the analyses. A total of 110 items did not have any topics, 82 of them belonging to research-article, 4 to editorial, 2 to article-commentary, 15 to image, and 7 to introduction to images.

These databases have been then used in the study in two ways. First, for the yearly analysis, they were split between the different years from 1991 to 2019. Second, for the cumulative case analysis, the databases were split in a way to cover the time period between the first year of publication of CHAOS (1991) and the considered year of interest. So, in this case, the year 1992 is represented by the data from 1991 to 1992, the year 1993 by the data from 1991 to 1993, and consequently the year 2019 represent the entire datasets covering all the years.

All the graphics were built using *ggplot2*[27] package in R[28], except the three network visualizations (Fig. S8 to S10) realized with Gephi software[29]. Networks were analyzed using *igraph*[30] package in R.

B. Method

TABLE II. Citations and topics in CHAOS. Total number of citations in Crossref ($Cit(CR)$), and ISI ($Cit(I)$) collected on 19th July 2020 and distributed according to the items' publication year. Temporal mean of citations in Crossref ($\langle Cit(CR)\rangle$), and ISI ($\langle Cit(I)\rangle$). Finally, the yearly evolution of the total number of topics (T), and the number of different topics (DT). The last row takes into account the whole period 1991-2019.

Year	$Cit(CR)$	$Cit(I)$	$\langle Cit(CR)\rangle$	$\langle Cit(I)\rangle$	T	DT
1991	1632	758	56.3	26.1	155	112
1992	1681	1412	60.0	50.4	158	106
1993	2047	1620	75.8	60.0	194	134
1994	1910	1851	73.5	71.2	198	117
1995	5154	5609	206.2	224.4	251	154
1996	1425	1546	59.4	64.4	184	139
1997	2782	2942	121.0	127.9	192	111
1998	3411	3579	155.0	162.7	315	173
1999	3510	3792	167.1	180.6	282	182
2000	3029	3179	151.5	159.0	255	154
2001	3208	3300	168.8	173.7	335	202
2002	3229	3287	179.4	182.6	344	190
2003	2927	3075	172.2	180.9	338	181
2004	2857	2870	178.6	179.4	359	209
2005	4339	4403	289.3	293.5	1457	575
2006	4033	4175	288.1	298.2	1654	534
2007	4159	4157	319.9	319.8	1712	584
2008	4782	4677	398.5	389.8	1912	555
2009	5703	5668	518.5	515.3	2015	609
2010	4302	4164	430.2	416.4	1990	648
2011	3747	3645	416.3	405.0	1910	603
2012	4266	4171	533.3	521.4	2439	655
2013	2833	2678	404.7	382.6	1838	599
2014	2228	2170	371.3	361.7	1779	596
2015	3744	3618	748.8	723.6	2953	794
2016	2581	2409	645.3	602.3	2805	725
2017	3080	2812	1026.7	937.3	3749	859
2018	2306	1898	1153.0	949.0	4223	863
2019	942	1022	942.0	1022.0	3437	773
1991-2019	91847	90487	-	-	39433	2375

We ordered the data according to relevant aspects that are used further in Sects. II C, III and IV. Thus, in Table S1, it is shown the evolving aspects of CHAOS, such as the total number of papers (NP), authors ($N(A)$), the number of different authors ($N(DA)$) and countries ($N(DC)$), the productivity ($P(X)$) and the degree of collaboration ($C(X)$) with respect to the authors and countries. The definition of $P(X)$ is given by $P(X) = \frac{NP}{N(X)}$, where X stands for A, DA or DC; while the degree of collaboration $C(X)$ is defined as $C(X) = \frac{N_m(X)}{NP}$, where $N_m(X)$ represents the number of papers with multiple X (authors: A or countries: C).

Other important data to characterize publication dynamics are the number of citations (Cit) according to the databases: Crossref (CR) and ISI (I), and also the topics of the articles (T). Citations were collected on 19th July 2020; they represent the number of citations that an item has accumulated till the collection date. To avoid the natural bias due to time, the temporal mean of citations of items published in a specific year i, $\langle Cit \rangle_i$ (see Eq. (1)), was calculated using the ratio of the number of total citations for the published items during this year (cumulative), $Cit_i^{(\text{total})}$, to the number of elapsed years since the publication year, i:

$$\langle Cit \rangle_i = \frac{Cit_i^{(\text{total})}}{2020 - i} \, . \tag{1}$$

Details of these quantities are represented in Table II.

C. Statistical aspects

A total of 5056 articles were published in CHAOS between 1991 and 2019, 95% of them being Research articles (Table S1). They were written by 15502 authors, from whom 8799 were different authors, belonging to 95 different countries (Table S1). The number of articles by year increased with time (Figs. 1, 2). Three periods can be distinguished. The first one, with less than 100 articles by year, lasted from 1991 to 2001; almost all the articles were Research-articles. The second period lasted from 2002 to 2014, with a number of articles being comprised between 100 and 200 a year (except a peak in 2012 with almost 250 articles). In this period, the Journal diversified the type of articles they published. For example, Image gallery and their introduction were present from 2004 to 2011. The last period started in 2015, with a substantial rise in the number of published articles (almost 500 in 2019). On the other hand, the number of issues changed from 4 to 12 per year in this last period. Fast tracks have started to be published in 2017. Focus issues have begun in 2005, and review articles in 2013 (Fig. 1).

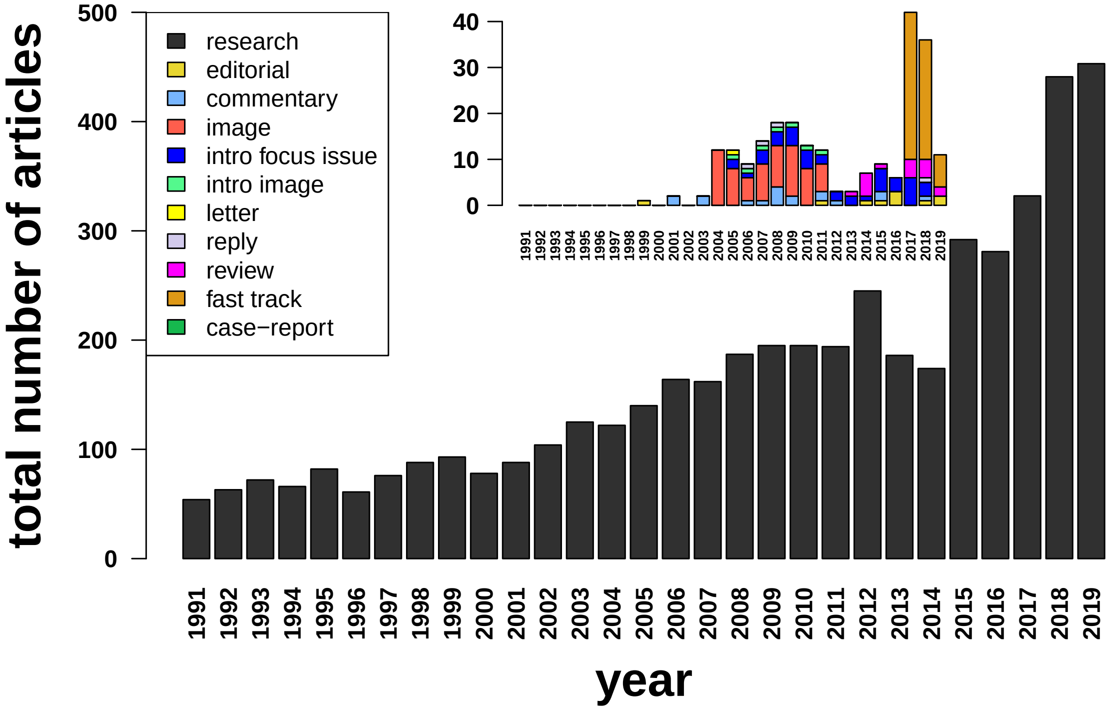

FIG. 1. Evolution of the total number of articles published yearly between 1991 and 2019 in CHAOS according to the type of article. As the number of research articles is far more numerous, the other types of articles (editorial, commentary, image, introduction to focus issue, introduction to image, letter, reply, review, fast track, and case report) are represented as an inset-plot with another scale.

The number of authors per article is comprised between 1 to 13 authors (with an exception of an article with 36 authors), the majority of the articles being written by 2 or 3 authors (Table III). In Fig. 3, we observe that the number of authors per article has increased between the beginning and the end of the observed period. This result is also reflected in the decrease of the authors' productivity and the increase of the collaboration degree by author during the same period (Fig. 2), showing that nowadays, the number of authors per article is higher.

Most of the articles were published by authors belonging to one country, knowing that up to six countries participated in some articles (Table III). The country contributing the most over the 29 years is the USA, participating in more than 20% of the publications, followed by China (10%) and Germany (9%) (Fig. 4). Over the whole period 1991-2019, the number of countries involved per article tend to increase (Fig. 3). This is also reflected in the rise of the collaboration degree

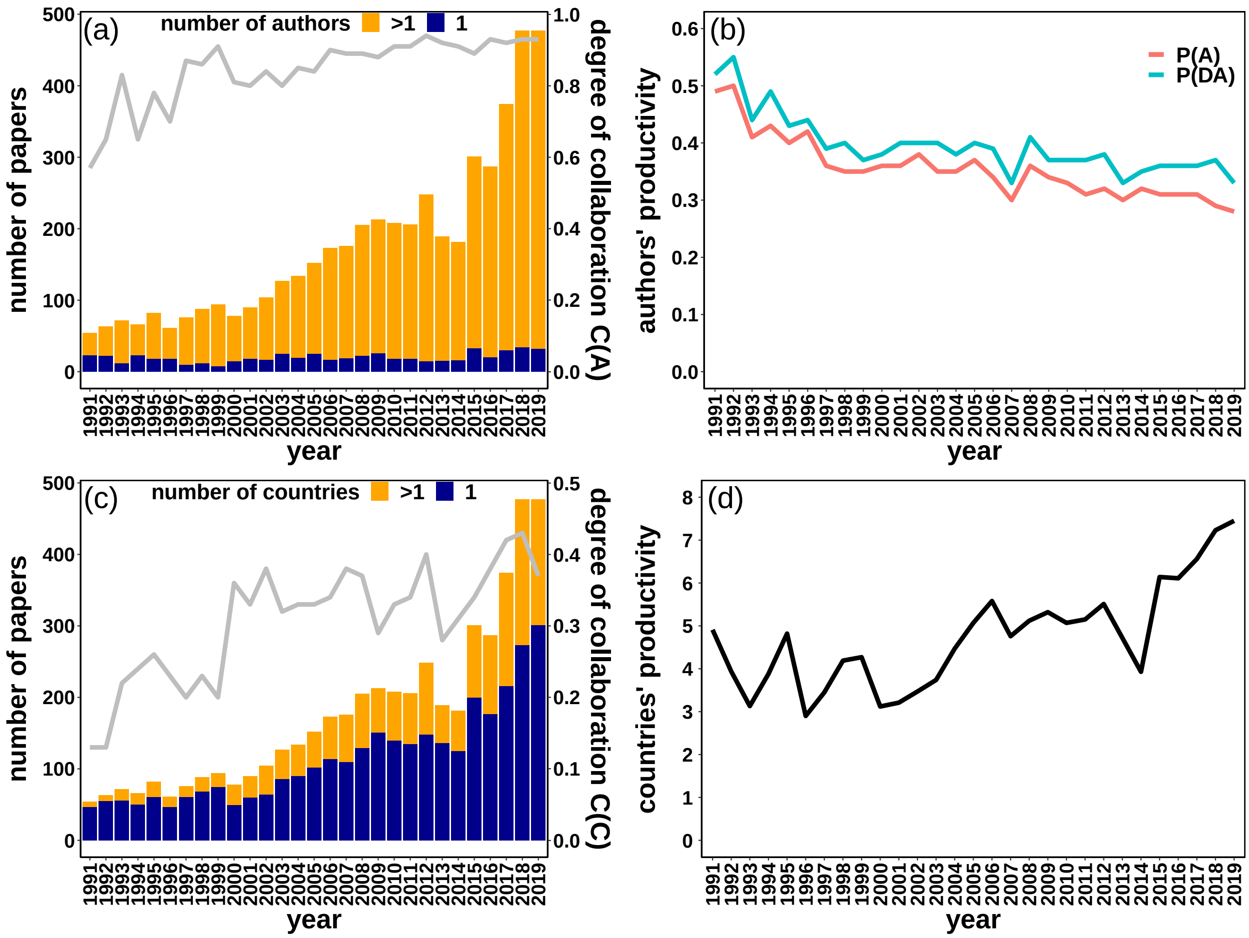

FIG. 2. Publication dynamics of CHAOS: number of papers published by one or more authors, and the corresponding degree of collaboration $C(A)$ (grey line) in (a), productivity of authors $P(A)$ and of different authors $P(DA)$ in (b), number of papers published by 1 or more countries, and the corresponding degree of collaboration $C(C)$ (grey line) in (c), and productivity of countries $P(DC)$ in (d).

by country between 1991 and 2019 (Fig. 2). The countries' productivity also increased over the years, meaning that the number of published articles increased more intensively than the number of countries participating in the publication.

A total of 2375 different topics have been detected over the whole period (Table II). A median of three topics per article was given during the period 1991-2004. From 2005, the median has increased to 10. This shift in the number of topics may reflect a change in CHAOS publication policies (Fig. 3). A total of 91,847 (Crossref) and 90,487 (ISI) citations were calculated for the whole period 1991-2019 (Table II). The citation numbers are similar between Crossref and ISI metrics, except for the first years (1991-1994). It is remarkable that using the temporal mean

TABLE III. Number of papers according to the number of participant authors and countries

Number of incidences	Author	Country
1	581	3327
2	1500	1315
3	1341	308
4	889	80
5	411	21
6	188	5
>6	146	0

criterion to avoid the natural bias due to the difference in time of the accumulated citations, we observe an increasing trend in the number of citations gathered by the cited items (Fig. 3).

III. ZIPF'S LAW

In 1926, Lotka reported a result related to scientometric aspects in which he found that the frequency distribution of scientific productivity (percentage of authors vs. number of citations) follows a power-law of the type $f(x) = x^{-\alpha}$, where $\alpha \approx 2$[31]. Later, in 1949, Zipf generalized the power laws' features to different types of human behavior[12]. Finally, a revaluation of the so-called Lotka's law for scientific productivity was carried out by MacRoberts in 1982[32]. Here, we analyze whether or not some dataset under study follows a power-law. In order to perform the above-mentioned analysis, rank-frequency distributions were obtained for 15 different datasets:

(a) number of publications by author, country, and topics (datasets 1, 2, 3),

(b) sum of Crossref citations by author, country, and topics (datasets 4, 5, 6),

(c) sum of ISI citations by author, country, and topics (datasets 7, 8, 9),

(d) sum of weights by author, and country (datasets 10, 11),

(e) h-index based on Crossref citations by author, and country (datasets 12, 13),

(f) h-index based on ISI citations by author, and country (datasets 14, 15),

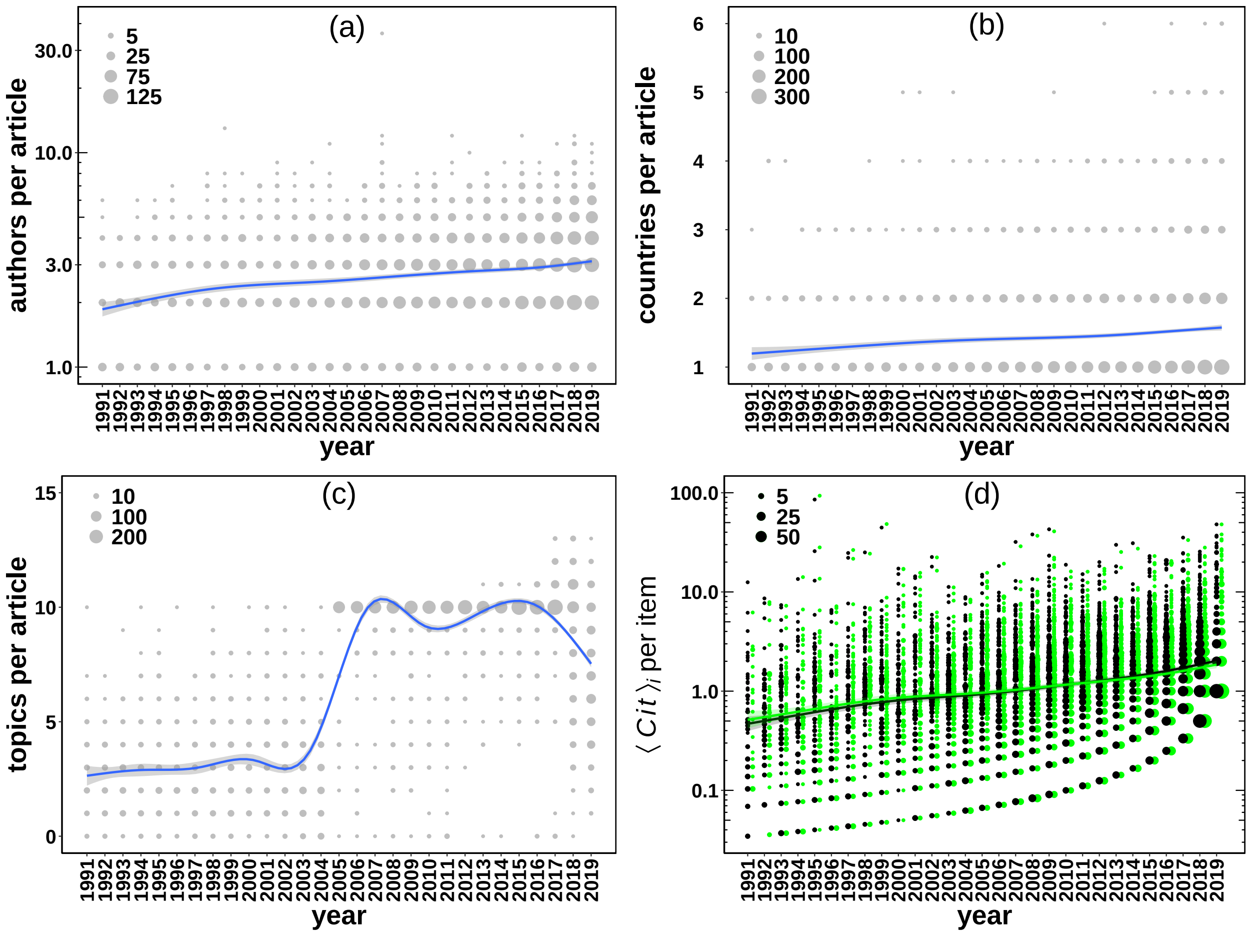

FIG. 3. Evolution over years of the number of authors (a), countries (b), and topics (c), and of the temporal mean of citations (d: black and green corresponding to Crossref and ISI, respectively, items with no citation are not represented). The bubble size is according to the number of published items found for the corresponding number of authors, countries, topics or citations. The global trend of each plot is represented by a generalized additive model (GAM) smoothing line.

These datasets were tested for power-law, lognormal, Poisson, and exponential distribution (except Poisson for continuous datasets) based on a fitting of the cumulative distribution function (CDF) using the method described in Clauset et al.[33] and implemented in the *poweRlaw* package of R[34]. The lower bound $x_{\min}$, i.e., the minimum value of x from which the distribution applies, and the parameters of each distribution were calculated using the Kolmogorov-Smirnov statistic: α for power-law, μ and σ for lognormal, and λ for Poisson and exponential distributions. Goodness-of-fit tests were performed via a bootstrapping procedure (2500 bootstraps) for all the distributions. To determine the plausibility of one distribution, we used the conservative criterion $p > 0.1$[33]. In

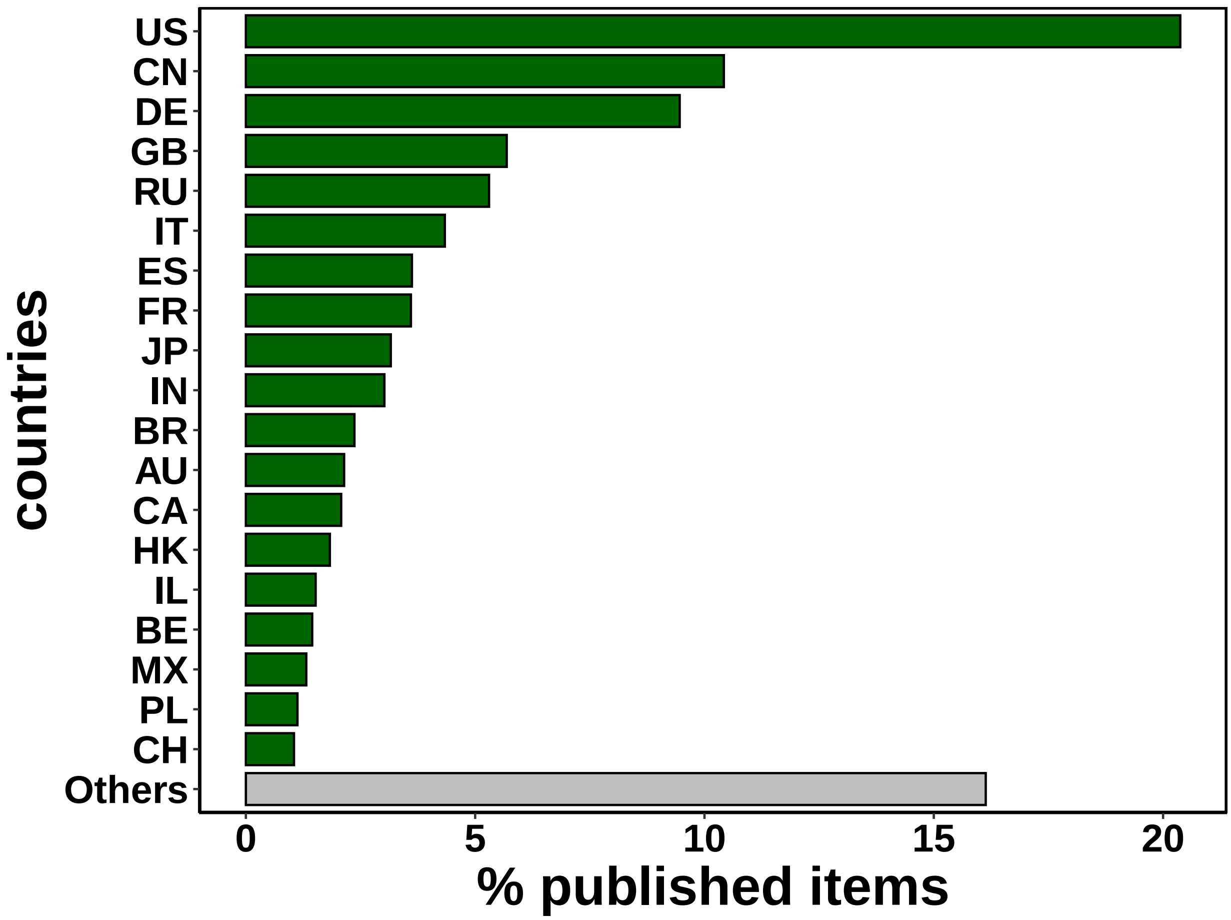

FIG. 4. Countries' publication between 1991 and 2019. The 19 countries with a contribution of at least 1% of the published items are represented in decreasing order (green), the remaining 76 countries being represented in Others.

Table IV, we observe that 7 distributions might follow a power-law, excluding models 1, 3, 8, 9, and 12 to 15.

When different distributions including the power-law were plausible for a dataset, the power-law was compared to alternative hypotheses via a likelihood ratio test[33]. For this comparison, the $x_{\min}$ value has to be equal for both distributions, and so, it was fixed as the $x_{\min}$ value found for the power-law. When $p \leq 0.1$, one distribution is better than the other, and the sign of the R ratio indicates which one ($R > 0$: power law is better, and $R < 0$: the other candidate distribution is better). In the case of $p > 0.1$, both distributions fit well and no conclusion can be made. Therefore, from the 7 datasets where the power law is plausible, in 6 cases the alternative distributions cannot be rejected (Lognormal in 4 datasets, Lognormal and exponential distributions

TABLE IV. Values of parameters and p-value of each distribution

dataset	power-law			lognormal				Poisson			exponential		
	p	x_{min}	α	p	x_{min}	μ	σ	p	x_{min}	λ	p	x_{min}	λ
1	0.000	1	2.46	**0.553**	1	-2.87	1.72	0.000	1	1.26	0.000	7	0.18
2	**0.540**	22	1.74	**0.120**	1	1.97	2.35	0.000	272	639.86	**0.180**	59	0.00
3	0.005	31	2.11	**0.586**	1	0.04	2.22	0.000	144	333.26	**0.180**	182	0.00
4	**0.879**	137	2.69	**0.789**	13	2.22	1.53	**0.543**	2461	2577.27	**0.626**	581	0.00
5	**0.513**	143	2.72	0.000	6	2.57	1.41	**0.749**	2659	2676.30	**0.586**	613	0.00
6	**0.358**	187	1.59	**0.718**	23	5.69	2.06	0.006	4873	12770.43	0.000	1	0.07
7	**0.274**	186	1.57	**0.784**	21	5.46	2.19	0.006	4709	12595.71	0.000	1	0.06
8	0.051	516	2.12	**0.112**	9	3.56	2.07	0.000	5425	9556.50	0.000	1	0.10
9	0.069	444	2.06	**0.326**	3	3.70	2.00	0.000	5150	8880.86	0.000	1	0.11
10	**0.814**	1	2.62	**0.611**	1	-1.55	1.28	NA	NA	NA	**0.615**	4	0.45
11	**0.156**	1	2.64	**0.639**	1	-1.60	1.28	NA	NA	NA	**0.906**	4	0.45
12	0.000	1	2.62	0.046	1	-1.55	1.28	0.003	9	9.89	0.032	4	0.45
13	0.000	1	2.64	0.046	1	-1.60	1.28	0.000	9	9.85	0.001	4	0.45
14	0.000	4	3.75	0.047	2	-0.15	0.94	0.004	9	9.89	0.034	8	0.16
15	0.000	19	1.73	0.043	0	1.71	2.33	0.000	9	9.85	0.002	256	0.00

in 2 datasets, see Table V). In dataset 5 (sum of Crossref citations by country), the power-law appears to be significantly better than all alternatives. The uncertainties of the parameters x_{min} and α of the plausible power-laws were determined by using 2500 bootstraps and computing the mean value and the standard deviation of the 2500 constructed surrogate data. These values are given in Table V. The plausibility that the sum of Crossref citations by country obeys a power-law distribution is related to the named Matthew effect which is also present in scientometric studies as stated by Perc[35].

IV. NETWORK ANALYSIS

The evolution of concepts arising from graph theory led to the notion of complex networks, which are of fundamental importance for science[36,37], due to the ubiquity of these structures in a

TABLE V. Comparison between power-law and other plausible distributions via a likelihood ratio test. R ratio, p-value, $x_{\min}$ and the result about which distribution is the most plausible are indicated. The uncertainty of the parameters ($x_{\min}$, α) of the power-law is given for each dataset.

dataset	comparison with	R ratio	p	$x_{\min}$	most plausible	uncertainty $x_{\min}$	uncertainty α
2	lognormal	-0.927	0.354	22	both	19.3 ± 17.7	1.7 ± 0.2
2	exponential	1.644	0.100	22	both		
4	lognormal	-0.360	0.719	137	both	124.5 ± 35.3	2.6 ± 0.2
4	Poisson	5.134	0.000	137	power		
4	exponential	4.089	0.000	137	power		
5	Poisson	4.819	0.000	143	power	141.2 ± 36.7	2.7 ± 0.1
5	exponential	4.152	0.000	143	power		
6	lognormal	-1.188	0.235	187	both	298.2 ± 477.9	1.6 ± 0.2
7	lognormal	-1.311	0.190	186	both	365.4 ± 648.2	1.7 ± 0.3
10	lognormal	0.070	0.944	4	both	2.3 ± 0.8	3.2 ± 0.3
10	exponential	1.674	0.094	4	power		
11	lognormal	-0.686	0.493	19	both	93.2 ± 1153.0	1.8 ± 0.3
11	exponential	1.235	0.217	19	both		

diversity of systems. It is precisely these networks, whose edges represent social connections, that are of interest for the description of collaborations[40] and used for scientometric studies. In this section, we analyze separately the adjacency matrices $\mathbf{A} = \big((a_{ij})\big)$ associated with the complex networks of authors, countries, and keywords (topics) of the articles. These items play the role of nodes, and the connections between them constitute the network's edges. The elements of the adjacency matrix a_{ij} with $i, j = 1, \ldots, N$, being N the number of nodes (authors, countries or topics). The nodes' relationship is considered from a dynamical system point of view, stated explicitly by Newman[39], and implicitly in other works[41], i.e., a_{ij} stands for the tie from j to i. Distinctions are made in the types of considered networks, as explained in the subsections below.

A. Unweighted and undirected case: diversity

The diversity of authors, countries, and topics present in CHAOS might be in a first approximation characterized by the simplest type of network, i.e., undirected and unweighted whose adjacency matrix elements are given by:

$$a_{ij} = \begin{cases} 1, & \text{if } i \text{ is linked to } j \\ 0, & \text{if } i \text{ is not linked to } j \text{ or } i = j \end{cases},$$

from which some basic indicators are determined, such as: (i) the number of edges and nodes, (ii) the number of isolated nodes, (iii) the number and size of subgraphs, i.e., disconnected fragments of the network, (iv) the degree, i.e., the number of links that a node has with other nodes[38], (v) the average of local clustering coefficients (LCC), i.e., the average of the LCC over all the nodes, representing the probability that two neighbors of a randomly selected node link to each other[38], (vi) the maximum of the betweenness centrality, which captures how much a given node is in-between others, measured by the number of shortest paths between two nodes passing through the target node[39], and (vii) the communities, i.e., set of nodes from a connected graph, for which the number of edges between them is greater than the number of edges linking them to the rest of the graph[40]. Given that the network's diversity is closely related to the number of nodes and their interrelations, Fig. 5 represents the evolution of diversity, where we choose the cumulative case corresponding to the number of nodes and edges for each feature of the nodes (authors, countries, and topics), and to the number and size of the subgraphs. The number of edges increases almost linearly with the number of nodes for authors (Fig. 5(a)), with an exponential trend for countries (Fig. 5(b)), and exponentially for topics (Fig. 5(c)). Numerous subgraphs (going from 29 in 1991 to 287 in 2019) are identified for authors each year; they are characterized by a mean size of less than five nodes (Fig. S1). In general, since 2003, the main subgraph defined by its size (number of nodes) has stood out from others (Fig. 5(d)), containing around 55 nodes as a mean (range: 10 to 244, median: 40). Nevertheless, it represents at most, only 19% of the total number of nodes (in 2018, where it has 244 nodes, Fig. S2), indicating that the authors' network is strongly fragmented with the presence of many cliques (i.e., complete network, where all nodes are connected to each other, and characterized by a unitary local clustering coefficient as shown in Fig. S3). For countries, in most of the years, all the nodes are connected, constituting a single component network (Fig. 5(e)). For topics, two periods are distinguishable: from 1991 to 2004, several subgraphs are present with one of them being much larger in size. After 2004, the main subgraph strongly increases in size with the consequent diminution in the number of subgraphs, and in 2005 and from 2011, a single component network is present (Fig. 5(f)). This observation can be related to the abrupt transition (2004-2005) of the number of topics attributed to each article, probably reflecting a change in the publishing policies of CHAOS.

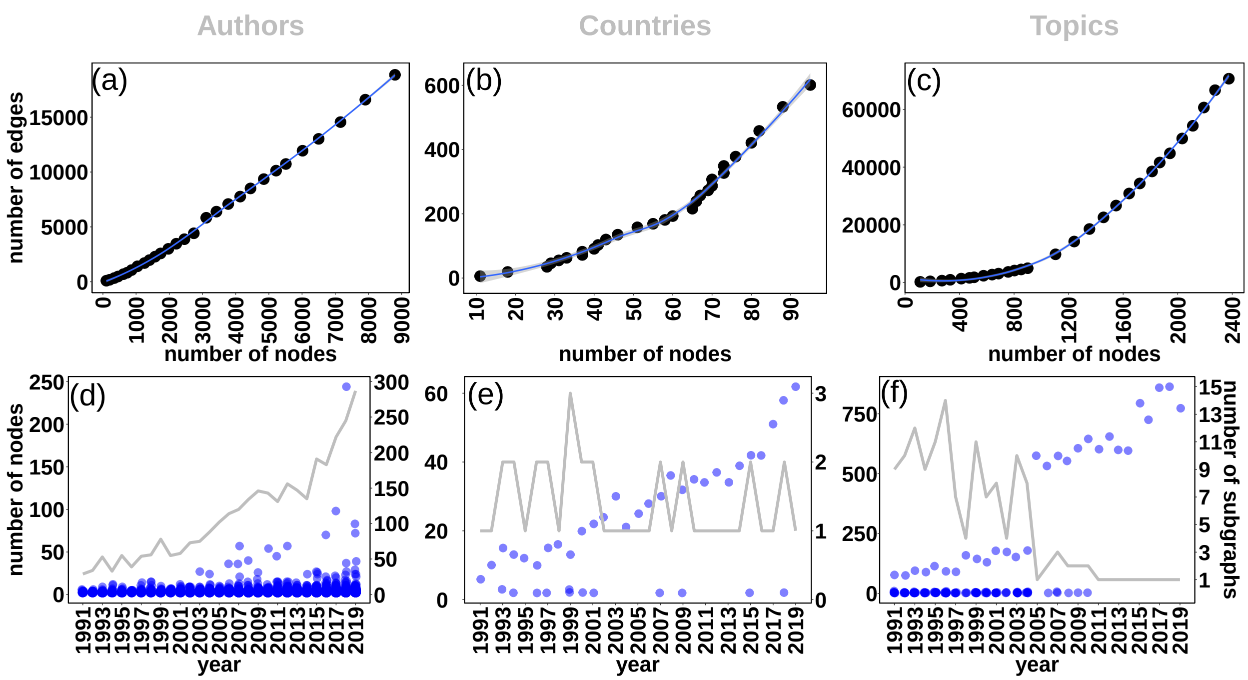

FIG. 5. Diversity evolution considering the cumulative case of the number of edges vs. nodes for the (a) authors, (b) countries, and (c) topics. Yearly number (grey line, left-y-axis) and size (blue points, right-y-axis) of the subgraphs for (d) authors, (e) countries, and (f) topics.

From the unweighted and undirected matrices analysis, it is interesting to note that the number of isolated nodes tends to diminish. Indeed, their fraction goes from 0.173 in 1991 to 0.016 in 2019 for authors, from 0.454 to 0.031 for countries, and from 0.027 to zero for topics (Fig. S4). That reflects the importance of belonging to a network, especially in the case of authors who tend to collaborate nationally and even internationally. This behavior is manifested as well from the statistical approach in Fig. 2(a). Probably this trend is accomplished in most of the scientific journals.

Another interesting quantity is the average local clustering coefficient that is obtained considering that nodes with degree zero and one are reported with zero transitivity. This coefficient tends to increase for the authors (0.56 to 0.83 from 1991 to 2019), has a slight tendency to increase for countries (0.19 to 0.46 from 1991 to 2019, but with very fluctuating values in between), and

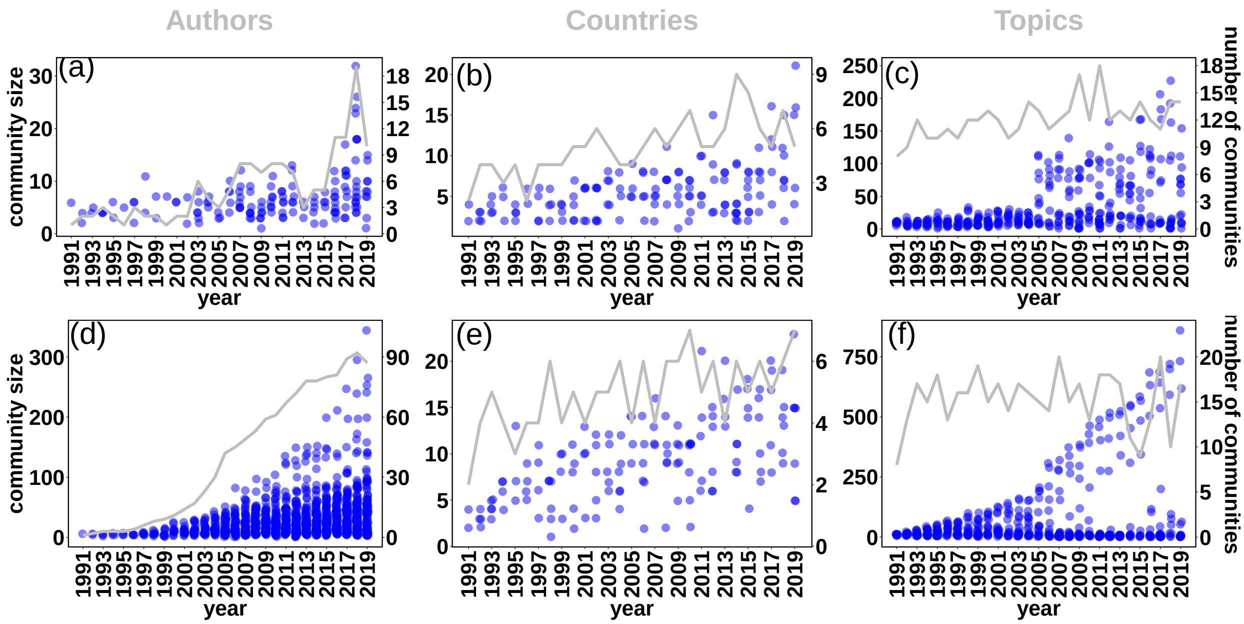

FIG. 6. Evolution of the number (grey line, left-y-axis) and size (blues points, right-y-axis) of the communities in the main subgraph (subgraph with the highest node number) for the (a, d) authors, (b, e) countries, and (c, f) topics. Non-cumulative (a, b, c) and cumulative (d, e, f) cases are considered.

remains constant for topics at around 0.73 ± 0.03 (Fig. S5).

Concerning the betweenness, we focus on its maximum along the years of existence of CHAOS: the three features (authors, countries, and topics) tend to increase, but in the case of authors and topics, this tendency intensifies from around 2004 (Fig. S6), indicating the presence of hubs (superconnectors or highly influential nodes[42]), which is corroborated by the representation of the nodes' degree distribution of the main subgraph (Fig. S7).

Finally, the presence of communities, i.e., cohesive groups of nodes which are more interconnected between themselves than with the rest of the graph[43], was investigated inside the main subgraph (the fragment of the network that has the highest number of interconnected nodes), for authors, countries, and topics, in the non-cumulative and cumulative cases (Fig. 6). Different algorithms have been developed to detect the communities inside a graph, widely explained in books devoted to networks[38,39,44]. We chose an algorithm based on statistical mechanics, concretely the spin-glass model and simulated annealing, for its performance[45], and implemented in *igraph*[30] package in R.

For authors, regarding the non-cumulative case (Fig. 6(a)), the number of the communities increases throughout the years, with the highest number in 2018, reflecting the main subgraph

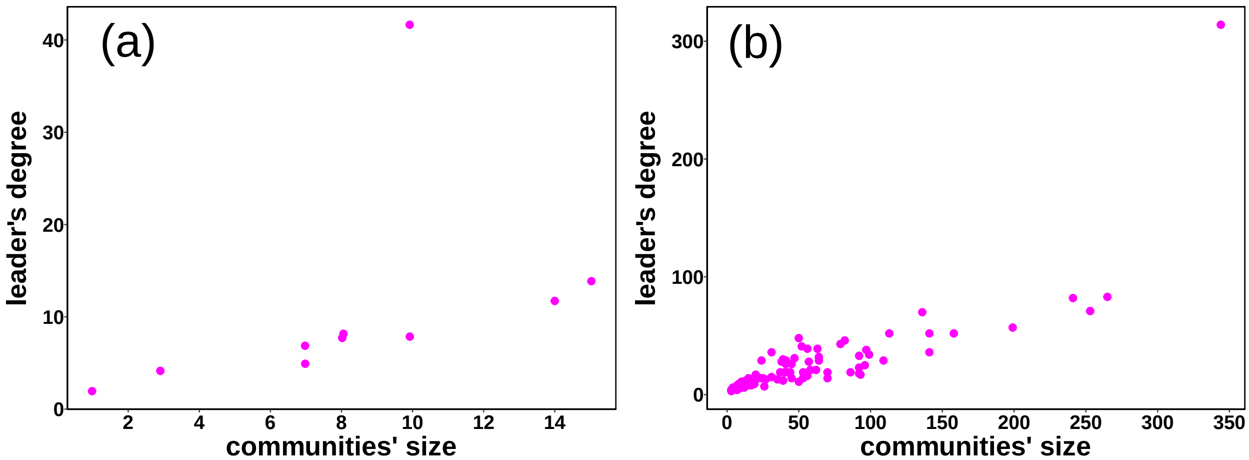

FIG. 7. Community description inside the main authors' subgraph for 2019 (a), and for the cumulative case 1991-2019 (b), where 10 and 87 communities were detected, respectively. The degree of the node representing the leader author is represented as a function of its community's size.

size observable on Fig. 5(d). A total of 10 communities was identified in 2019, all of them led by an author whose degree is less than 15, except for the presence of a supernode with degree 42 (Fig. 7(a)). In the cumulative case, the size of the communities increases strongly throughout the years, with some of them being larger with more than 100 nodes, and also with the biggest community having more than 300 nodes (Fig. 6(d)). For the period 1991-2019, 87 communities were identified, whose size varies between one and 344 authors. The degree of the leading author of each of the communities is less than 100, except for one community whose leading author has a degree equal to 314 suggesting a supernode (Fig. 7(b)).

For countries, the yearly number and size of communities tend to remain constant, with just a slight increase in the size of the main communities since 2016 (Fig. 6(b)). In the cumulative case, the community size has also the tendency to increase but is never larger than 23 nodes (Fig. 6(e)).

In the case of countries, it is important to visualize the communities of the main subgraph onto a geographical map. In Fig. 8, we took as an example the communities for 2019 and also for the cumulative period 1991-2019. The communities are characterized by their size and the nodes degree. In which follows, we describe each community giving its size (N_i), and the main nodes with their degree (k). In 2019, five communities are distinguished: (i) $N_1 = 21$, USA ($k = 25$) and Germany ($k = 24$); (ii) $N_2 = 16$, Italy ($k = 22$) and China ($k = 21$); (iii) $N_3 = 15$, Turkey ($k = 16$); (iv) $N_4 = 6$, Iran ($k = 16$); and (v) $N_5 = 4$, Cameroon ($k = 7$). For the period 1991-2019, seven communities are distinguished: (i) $N_1 = 15$, USA ($k = 57$); (ii) $N_2 = 15$, Germany

($k = 52$); (iii) $N_3 = 23$, United Kingdom ($k = 44$); (iv) $N_4 = 9$, Spain ($k = 43$); (v) $N_5 = 15$, India ($k = 30$); (vi) $N_6 = 5$, Iran ($k = 24$); and (vii) $N_7 = 5$, Cameroon ($k = 13$). In both cases, all the communities expand on different continents.

For topics, two periods are distinguishable: before 2005, the communities are numerous and small, and since 2005, the size of the communities increases (Fig. 6(c)). In the cumulative case, 2005 is also a pivotal year, with the emergence of dominant communities with a size being much larger than the other ones (Fig. 6(e)). In Table VI, the characteristics of the communities found in the main subgraph in 2019 and in 1991-2019 are given. A total of 14 communities were detected for 2019, 50% of them have a size greater than 60 nodes. The leading topics with the highest degree are: Dynamical systems (339) and Stochastic processes (213). For the period 1991-2019, 17 communities were discovered, three of them have a very huge size (860, 732, and 618 nodes respectively), and 12 of them have a very small size with less than 10 nodes. The topics with the highest degree are also Dynamical systems (1324), and Stochastic processes (1057). As in the other cases, the presence of hubs is evident.

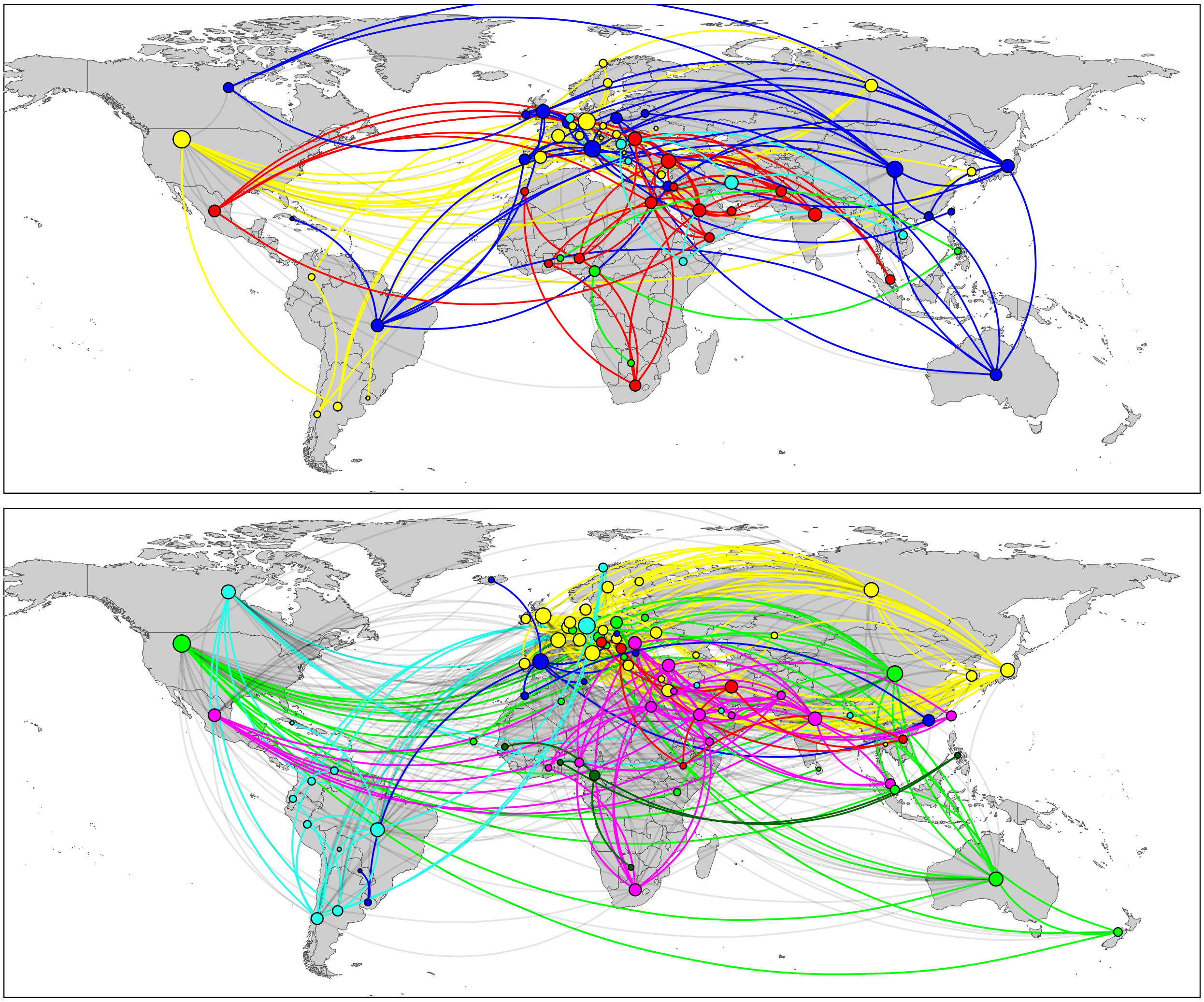

FIG. 8. Visualization of the communities inside the main subgraph for the year 2019 (top) and the entire period 1991-2019 (bottom). Nodes of the same color belong to one community. The links inside a community group are painted with the same color as the nodes, the links between communities being in grey. The size of the nodes increases with its degree. In 2019, 5 communities detected, with as country leaders: USA/Germany (yellow), Italy/China (dark blue), Turkey (red), Iran (light blue), and Cameroon (light green). In 1991-2019, 7 communities detected, with as country leaders: USA (light green), Germany (light blue), United Kingdom (yellow), Spain (dark blue), India (magenta), Iran (red), and Cameroon (dark green).

TABLE VI. Community description inside the main topics' subgraph for 2019 (left) and 1991-2019 (right). The number of communities (Comm), the leading topic of each community and its degree (deg), and the size of each community are given. Communities are ordered by the degree of their leading topic. Alg.: Algorithms, PDE: partial differential equations.

Comm	leading topic	deg	size
1	Dynamical systems	339	77
2	Stochastic processes	213	154
3	Mathematical modeling	195	94
4	Network theory	175	111
5	Chaotic dynamics	167	68
6	Nonlinear systems	164	68
7	Chaotic systems	131	20
8	PDE	100	76
9	Integral transforms	58	22
10	Alg. and data structure	53	9
11	Social science	48	55
12	Integral calculus	23	10
13	Mechanical stress	17	8
14	Digital circuits	7	1
-	-	-	-
-	-	-	-
-	-	-	-

Comm	leading topic	deg	size
1	Dynamical systems	1324	860
2	Stochastic processes	1057	63
3	Phase space methods	966	618
4	Signal processing	875	732
5	Non linear dynamics	719	53
6	Electric currents	61	9
7	Stereoscopy	37	8
8	Dielectric materials	32	5
9	Mean field potentials	19	4
10	3D printing	15	4
11	Equilibrium chemistry	10	4
12	Biomolecular structure	9	3
13	Photon absorption	9	1
14	Hall effect	7	3
15	Acousto-optics	6	1
16	Electric generators	5	1
17	Hertz' law	2	1

B. Weighted and undirected case: collaboration

Now, we use a more refined description of a network by including weighted links, indicating the number of publications that the authors, and countries have together, and characterized by a weighted and symmetric adjacency matrix whose elements are expressed by:

$$a_{ij} = \begin{cases} a_{ij} \in \mathbb{N}, \text{ if } i \text{ is linked to } j \\ 0, \text{ if } i \text{ is not linked to } j \text{ or } i = j \end{cases} .$$

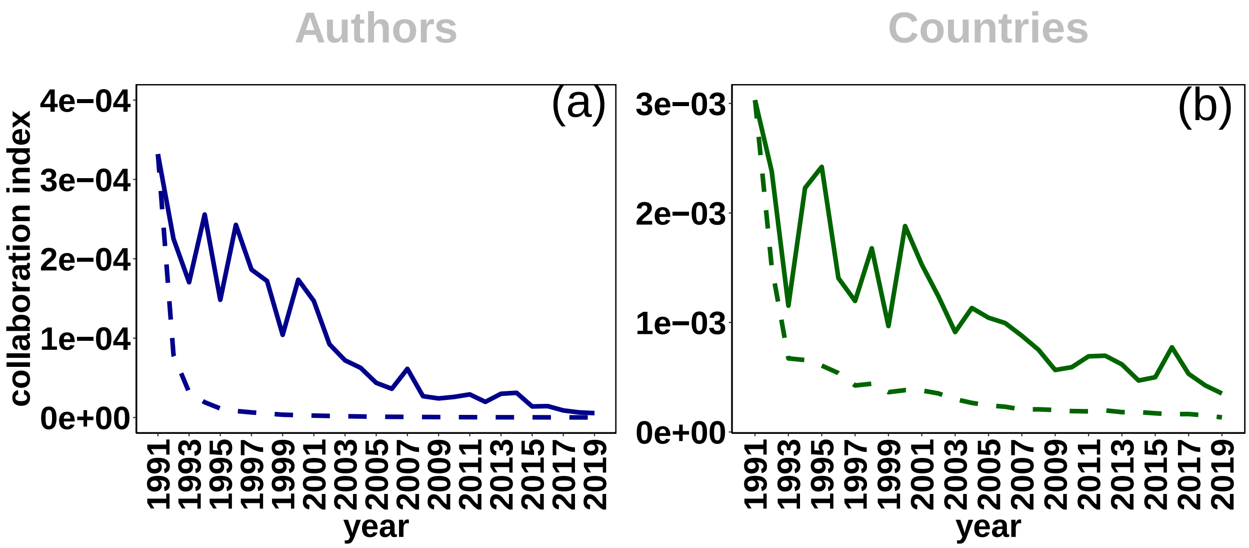

FIG. 9. Collaboration index obtained from Eq. (2) evolution considering the (a) authors, and (b) countries. Cumulative (dashed line) and non cumulative (solid line) cases are considered.

From this matrix, it is possible to define a general collaboration index $C(X)$, which is computed for the yearly and cumulative cases. The collaboration index is defined as the quotient of the matrix elements sum and a normalization factor, which depends on the number of papers NP and the number of the considered variable X (authors, or countries), and it is given by:

$$C(X) = \frac{\sum_{i=1}^{N(X)} \sum_{j=1}^{N(X)} a_{ij}}{N(X)(N(X)-1)NP} . \tag{2}$$

being $N(X)$ the number of authors or countries that published in CHAOS. It is clear that the validity interval is $0 \leq C(X) \leq 1$.

As shown in Fig. 9, the collaboration index tends to decrease with some fluctuations for the yearly case both for authors and countries. Instead, for the cumulative case, the index decreases abruptly and continuously during the first years, and afterward, there is a kind of stabilization for $C(X)$ with quite small values. The latter is reasonable because the number of papers rises yearly at an increasing rate especially from 2015 (see Fig. 1). The index is much less for the authors than for the countries because of the larger number of authors compared to the countries.

C. Weighted and directed case: influence

Unlike the previous cases, in which the article weights were not taken into account, we now consider authors and countries weights computed as explained in Table I. Therefore, asymmetric

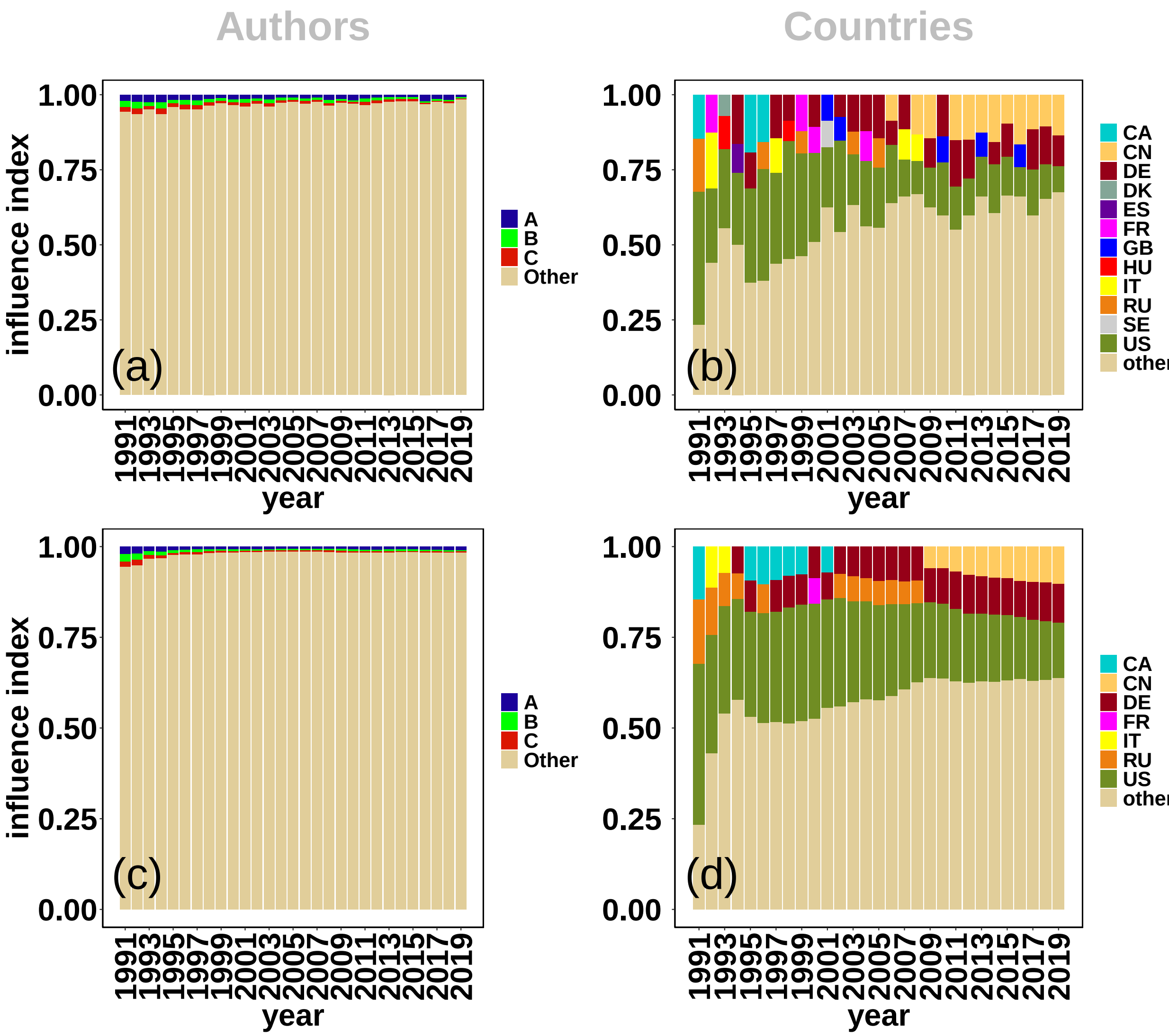

FIG. 10. Influence index (obtained from Eq. (3)) evolution of the first three highest ranked nodes, the other nodes being assembled in the "other" group. Authors (non cumulative (a) and cumulative (c) cases), as well as the countries (non cumulative (b) and cumulative (d) cases) were considered. For authors, the first three highest ranks are noted A, B, C, A being the highest and C the third ranked one. Thus, the letter does not correspond to a unique author, there is a change over the years. For countries, the three highest ranked countries are identified: CA: Canada, CN: China, DE: Germany, DK: Denmark, ES: Spain, FR: France, GB: United Kingdom, HU: Hungary, IT: Italy, RU: Russian Federation, SE: Sweden, US: United States of America.

adjacency matrices are obtained whose elements are given by:

$$a_{ij} = \begin{cases} a_{ij} \in \mathbb{R}_+, \text{ any number in the interval } (0, \infty) \\ 0, \text{ if } i \text{ is not linked to } j \text{ or } i = j \end{cases},$$

where the column corresponding to a given feature X (author or country) indicates the participation of X in joint publications with the author or country corresponding to each row.

Based on this type of network, we can calculate the influence of a node on the nodes with which it collaborates. Thus, an indicator of the j-th node influence can be defined as

$$I_j = \frac{\sum_{i=1}^{N(X)} a_{ij}}{\sum_{i=1}^{N(X)} \sum_{j=1}^{N(X)} a_{ij}} . \tag{3}$$

For the yearly case of authors (Fig. 10(a)), during the first years, the three most influential authors have an influence index within CHAOS of around 0.02 (0.06 for the three together), whereas in 2019, they only had indexes of 0.0066, 0.0052, and 0.0044, making a sum of 0.0162. These results show that due to the growing number of authors, a particular author's influence is negligible. For the cumulative case (Fig. 10(c)), this trend is even more visible, since the cumulative number of authors is significantly higher. The sum of the cumulative influence index for the three most influential authors is 0.0146. However, it is remarkable that the most influential author in the entire history of CHAOS has an index of 0.0096, significantly higher than 0.0032 and 0.0028, corresponding to the second and third most influential authors.

For the yearly case of countries (Fig. 10(b)), there were countries with a huge influence during the first years. Thus, in 1991, the indexes for the most influential countries were 0.44, 0.18, and 0.15 (USA, Russia, and Canada), representing a total influence of 0.77, while in 2019, these indexes fell to 0.14, 0.10, and 0.09 (China, Germany, and USA), which is equivalent to a total influence of 0.33. The latter is because contributions from other countries increased considerably. It is also observed that the ranking of the most influential countries changes year after year, although one of the countries (USA) has always been in one of the top three positions. For the cumulative case (Fig. 10(d)), a significant decrease is observed in the first years in the influence of the three best-placed countries in the ranking. Subsequently, there is a stabilization trend in the sum of the three most influential countries, which is around a third of the total influence.

On the other hand, it is also seen that the USA has been the most influential country in the entire history of CHAOS, that Germany entered the ranking of the three most influential in 1994 and remained uninterruptedly since 1997. Finally, China entered the ranking in 2009 with an increasing trend in influence in CHAOS history that allowed its influence index to grow from 0.06 to 0.10 in the last 11 years. Throughout CHAOS history, the most influential countries are the USA, Germany, and China, with indexes in 2019 of 0.15, 0.11, and 0.10, respectively.

D. General case with weighted nodes: productivity

In order to characterize the productivity, we consider both the specific weight of the participation of each node in the yearly total weight of CHAOS, and the cumulative participation throughout its history. An asymmetric adjacency matrix whose elements are given by:

$$a_{ij} = \begin{cases} a_{ij} \in \mathbb{R}_+^* \text{ any number in the interval } [0, \infty) \\ 0, \text{ if } i \text{ is not linked to } j \end{cases},$$

characterizes participation, since the nodes' contribution in each paper can vary according to the number of participants (authors or countries). Furthermore, there will be elements on the diagonal since, on certain occasions, the documents are only written by a single author or by authors belonging to a single country. With this adjacency matrix, it is possible to define several concepts that will be characterized by productivity indicators as explained below:

CHAOS' general production: The CHAOS' production in a certain time interval (for authors or countries) is defined as the quotient of the production for a time interval l and the sum of the yearly production along the entire CHAOS' history (1990-2019):

$$P^{(l)}(X) = \frac{\sum_{i=1}^{N(X)} \sum_{j=1}^{N(X)} a_{ij}^{(l)}}{\sum_{k=1990}^{2019} \sum_{i=1}^{N(X)} \sum_{j=1}^{N(X)} a_{ij}^{(k)}}\,, \tag{4}$$

where X stands for authors or countries.

Individual production of articles: The individual production in time for authors or countries is defined as the quotient of the sum of the individual productions in time (trace of the matrix corresponding to a specific time interval) and the total production during the time interval (sum of the adjacency matrix elements):

$$P_{\text{ind}}^{(l)} = \frac{\sum_{i=1}^{N(X)} a_{ii}^{(l)}}{\sum_{i=1}^{NC} \sum_{j=1}^{N(X)} a_{ij}^{(l)}}\,. \tag{5}$$

The collaborative production: Is the complement of the individual production:

$$P_{\text{col}}^{(l)} = 1 - P_{\text{ind}}^{(l)}\,. \tag{6}$$

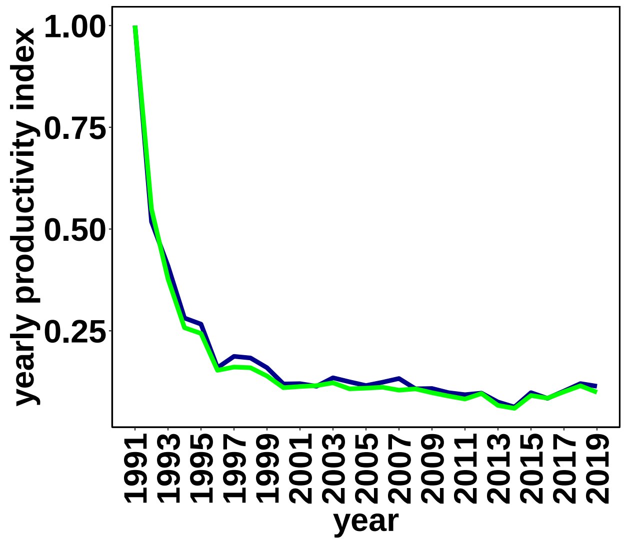

FIG. 11. Yearly productivity index evolution for authors (blue) and countries (green), obtained from Eq. (4).

The productivity per author or country j: Defined as the quotient of the sum of column j of the adjacency matrix corresponding to the author or country and the sum of all the elements of the adjacency matrix:

$$P_j = \frac{\sum_{i=1}^{N(X)} a_{ij}}{\sum_{i=1}^{N(X)} \sum_{j=1}^{N(X)} a_{ij}} \,. \tag{7}$$

The yearly productivity index decreases similarly for authors and countries from 1 to around 0.15, with an abruptly fall during the six first years, followed by a slow diminution (Fig. 11). The individual and collaborative productivity indexes are shown in Fig. 12. For authors, the individual productivity index had a high value of 0.29 during the first year of publication of CHAOS, and then decreased over time with fluctuations to reach 0.02 in 2019 (Fig. 12(a)). In the cumulative case, this decrease was smoother (Fig. 12(c)) and goes from 0.29 to 0.05 over the period. For countries, the individual productivity index is higher than for authors, as the number of countries is much smaller than the number of authors. Nevertheless, a similar decrease in this index is observed. The yearly individual productivity index fell from 0.85 to 0.55 with fluctuations (Fig. 12(b)). In the cumulative case, the decrease was smoother from 0.85 to 0.59 (Fig. 12(d)).

The productivity per author and per country is shown in Fig. 13. For authors, the three authors with the highest productivity demonstrated an individual productivity of around 0.02-0.03 each in 1991. This productivity decreased over time to reach 0.005 each in 2019.

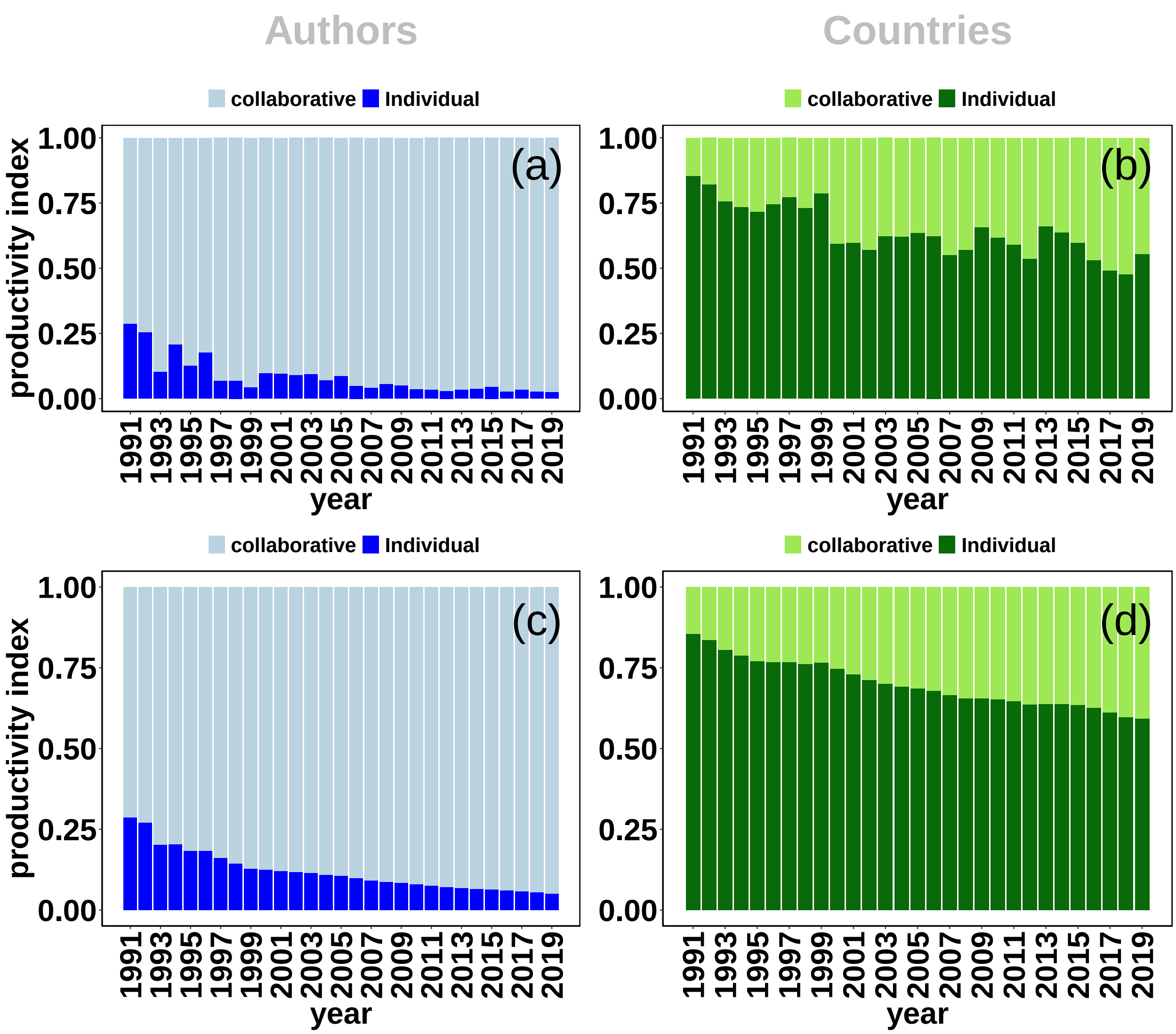

FIG. 12. Productivity index evolution for authors (non cumulative (a) and cumulative (c) cases), as well for countries (non cumulative (b) and cumulative (d) cases). The individual and collaborative productivity, obtained from Eqs. (5) and (6) respectively, are distinguished in each case.

In the cumulative case, the individual productivity decreased smoothly from 0.02-0.03 each in 1991 to reach a value of 0.009 for the most productive author and 0.002 for the second and third ones. Concerning the countries, it is interesting to note that during the first year, the first three most productive countries were the USA, Russia, and Canada, accumulating together an index of 0.88. In 1992, this individual productivity of the first three countries decreased, being 0.56, showing that since the second year of publication, CHAOS succeeds in attracting authors from more countries. In 2019, the indexes of the first three most productive countries sum together 0.44. The USA was always one of the three countries with the most intense productivity, in most of the cases occupying the first place in the ranking. China appeared from 2006, exhibiting increasing productivity throughout the successive years until reaching the first place of the ranking in the last years. Germany is the third country that comes regularly in the three most productive countries since 1993, being generally in the third place of the productivity ranking. In the cumulative case, this tendency observed in the yearly case is strengthened: in 1991 to 1991-92, the USA, Russia, and Canada were the most productive countries. Along the CHAOS' history, the USA is the most productive country, and until 2001, Russia kept the second place, but with a decreasing productivity. In 1993, Germany appeared and increased slowly its productivity occupying the second place in the global ranking from 2006 to 2011, when China emerged as the second most productive country throughout the CHAOS' history.

As an illustration, the weighted and directed network of authors obtained for the years 1991 and 2019 are shown in Fig. S8. We can appreciate the increase in diversity (higher number of authors in 2019), and also the presence of various single components. The weighted and directed network of countries for the cumulative case 1991-2019 is illustrated in Fig. S9. We can appreciate the size of the main subgraph, with only few countries having not linked to the other ones. We can also observed the high number of self-loops, showing the number of articles published by the country alone, and the presence of the seven communities. Finally, on Fig. S10, there is an illustration of the weighted and directed network of topics, for the cumulative case 1991-2019. The most important topics in terms of degree number are labeled.

V. CONCLUSIONS AND PERSPECTIVES

We performed a detailed scientometric analysis of CHAOS, starting with some standard descriptive statistics related to traditional bibliometric studies. It allowed us to analyze the CHAOS' publications evolution from its birth in 1991 to 2019. It is remarkable how the Journal grows, with around 50 articles per year published by 11 countries in its inaugural year to almost 500 in 2019 implicating 64 countries. The Journal also diversified its type of articles on the overall period, even if 95% are research articles. The increase in the temporal mean of citations also shows the rise of the Journal's impact in its field.

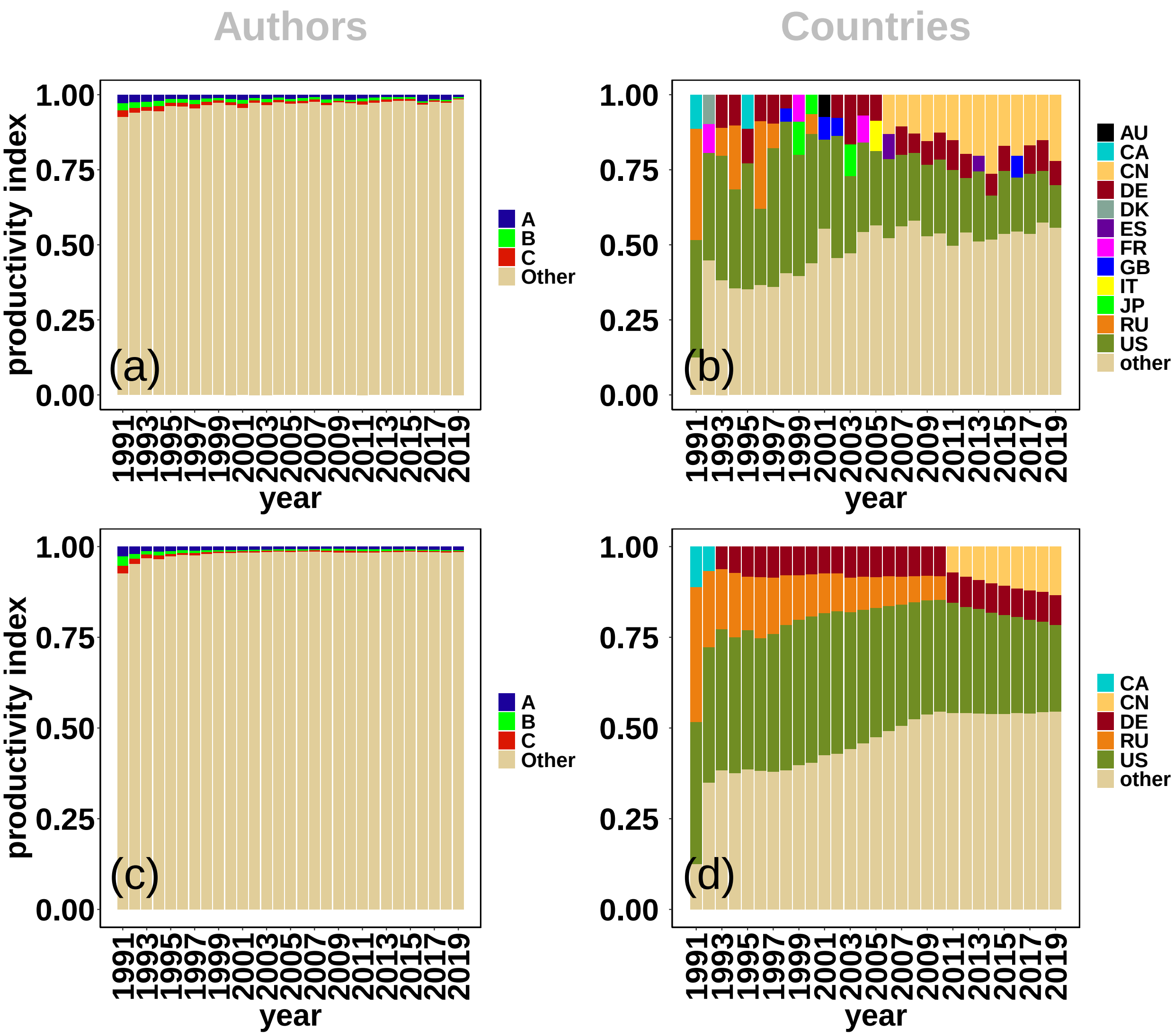

FIG. 13. Productivity per author (a, c) or country (b, d) index evolution for non cumulative (a, b) and cumulative (c, d) cases, obtained from Eq. (7). The three highest ranked individuals are represented, the rest being together in the Other group. The first three highest ranked authors are noted A, B, C, A being the highest and C the third ranked one. Thus, the letter does not correspond to a unique author, there is a change over the years. For countries, the three highest ranked countries are identified: AU: Australia, CA: Canada, CN: China, DE: Germany, DK: Denmark, ES: Spain, FR: France, GB: United Kingdom, IT: Italy, JP: Japan, RU: Russian Federation, US: United States of America.

Next, we verified the accomplishment or not of power-law in the rank-frequency distributions of the number of publications, the sum of Crossref and ISI citations, and the corresponding h-indexes by author and country. We have also considered the topics and citations for the analysis.

We found that the sum of Crossref citations by country is the only case where the power-law is the only possible distribution.

Unlike typical approaches, we then introduced new scientometric indexes from the underlying networks for authors, countries, and topics. The new concepts stated above depend on the structure and features of the networks. Thus, we used firstly the simplest case consisting of undirected and unweighted networks. It enabled us to describe aspects related to the diversity via the computation of the number of nodes and edges and the number and size of the subgraphs and quantities such as the node degree, clustering coefficient, and the maximum betweenness. In all cases, the number of nodes escalates, reflecting the enormously increasing diversity all through the Journal's history. Even though the networks' structure is quite different for authors (considerably fragmented, with the main subgraph including less than 20% of the total number of nodes), and for countries and topics (with almost all the nodes being part of the main subgraph), we found that in all cases, the number of nodes escalates, reflecting the strongly increasing in the diversity all through the Journal's history. We also determined the communities for authors and countries that gave us a first approach to visualize the collaboration. It is important to note that CHAOS attracts authors of all the continents and that the communities' study exhibits robust collaboration networks between countries or even different continents. The community analysis complemented by the computation of the maximum betweenness and the node degree distribution also shows the important presence and role of hubs. The collaboration strengthening is also reflected by the decline in the number of isolated nodes in the different networks. The analysis of the communities of topics highlights the presence of supernodes, being the main ones: Dynamical systems and Stochastics processes both in 2019 and in the interval 1991-2019. Subsequently, we considered the case of link-weighted and undirected networks, which permitted us to describe the collaboration between authors and countries, employing an index, which has small values both for authors and countries and decreases throughout the CHAOS' history due to the increasing number of authors and countries occurring each year. Another index describes the influence of authors and countries also defined in terms of the networks' features; in this case, considering link-weighted and directed networks. Our analysis uncovered that the most influential authors have less influence over time, which is related to the authors' increasing diversity. On the other hand, we determined the evolution of the most influential countries, noting among other aspects that the USA is the most influential country in the CHAOS' history followed by Germany and China, the last one having a strong tendency to increase its influence in the 11 previous years. Finally, considering the most general case, in which the nodes are also weighted, we study the productivity of authors and countries, introducing indicators such as the CHAOS' general production, the individual production, and its complement called the collaborative production of articles, considering authors and countries, and at the end, we computed the productivity per author and country. Our productivity results are correlated with those of influence, exhibiting, in particular, the supremacy in productivity of the USA in the complete CHAOS' history, but with the firm productivity of China in the last years, positioning this country as the most productive in the previous seven years.

It is noteworthy to highlight that the used metrics in statistics and complex networks are conceived differently. For the statistical case, simple quantities such as the number of authors (countries), the number of papers with single or multiple authors (countries) are enough to define the concepts related to collaboration and productivity. On the contrary, in the complex networks approach, the mentioned metrics depend on the well-defined nodes and edges that are clearly identified allowing us to define more metrics, all in terms of the elements of the adjacency matrix. Hence, contrary to scientometrics evaluation with descriptive statistics, the analysis with complex networks allows a better estimation of characteristics linked to collaboration and productivity. Moreover, the important concept of influence is introduced. The study with complex networks has more significant potential, and advantages for performing more objective comparisons which lie in the specification of weights to articles and the role of authors or countries in each article. The indexes found using complex networks are normalized, which would allow better comparisons of collaboration and productivity in different systems, namely the journals to be compared.

An important comment about the results deals with the robustness of the system in relationship with the influence of the nodes and the individual productivity. The fact that the results show small values of influence and individual productivity is an indicator of the journal's robustness because the extraction of the most influential and/or most productive nodes would not affect the network structure significantly. On the contrary, in some journals of smaller scope, there sometimes appear super influential and productive nodes. The above-mentioned feature is threatening because such nodes' extraction could strongly affect the system with possible negative consequences[46]. A new indicator related to the importance of a node that might be named as the preponderance of a node in the network could be estimated combining its influence and productivity.

The last issue to emphasize is that a thorough disambiguation process has been performed to identify the authors correctly for the analysis, with the consequent withdrawal of biases and misleading results.

We expect to extend this type of analysis in other contexts of scientometrics and introduce new indicators that might consider the concepts of multilayer networks.

SUPPLEMENTARY MATERIAL

The supplementary material is composed of Table S1, where the most important statistical data are presented, and Figs. S1–S10, which complement the explanation of results obtained in Sect. IV A, related to the evolution of networks features such as the size of the subgraphs, the local clustering coefficient, the fraction of the isolated nodes, the maximum betweenness, the degree of the nodes, and the networks visualization.

ACKNOWLEDGMENTS

We acknowledge CHAOS for allowing us to use its database. J.K. was supported by the Russian Ministry of Science and Education Agreement No. 075-15-2020-808.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] J. Mingers, and L. Leydesdorff, "A review of theory and practice in scientometrics," Eur. J. Oper. Res. **246**, 1–19 (2015).

[2] J. Tague-Sutcliffe, "An introduction to informetrics," Inform. Process. Manage. **28**, 1–3 (1992),

[3] P. Vinkler, *The evaluation of research by scientometric indicators*, (Chandos Publishing, Oxford, 2010).

[4] F. Cantú-Ortiz, *Research analytics: boosting university productivity and competitiveness through scientometrics* (CRC Press. Taylor & Francis Group, Boca Raton, 2017).

[5] P. Vinkler, "An attempt for defining some basic categories of scientometrics and classifying the indicators of evaluative scientometric," Scientometrics. **50**, 539–544 (2001).

[6]A. Moonghali, N. Alijani, A. Karami, and A. Khasseh, "Scientometric analysis of the scientometric literature," Int. J. Inf. Sci. Manage. **9**, 19–31 (2011).

[7]H. Padhy, P. Mishra, and S. Behera, "Bibliometric analysis tool - A review," Glob. J. Eng. Sci. Res. Manage. **6**, 1–7 (2019).

[8]L. Bornmann, and W. Marx, "Methods for the generation of normalized citation impact scores in bibliometrics: Which method best reflects the judgements of experts?," J. Informtr. **9**, 408–418 (2015).

[9]F. Galligan, and S. Dyas-Correia, "Altmetrics: rethinking the way we measure," Serials Rev. **39**, 56–61 (2013).

[10]M.C. Yu, Y.C. Wu, W. Alhalabi, H.Y. Kao, and W.H. Wu, "ResearchGate: an effective altmetric indicator for active researchers?," Comput. Hum. Behav. **55**, 1001–1006 (2016).

[11]J. Ortega, "Relationship between altmetric and bibliometric indicators across academic social sites: The case of CSIC's members," J. Informtr. **9**, 39–49 (2015).

[12]G. Zipf, 1949, *Human behaviour and the principle of least effort: an introduction to human ecology*, (Addison-Wesley, Cambridge, 1949).

[13]J. Hirsch, "An index to quantify an individual's scientific research output," P. Natl. Acad. Sci. USA. **102**, 16569 (2005).

[14]F. Guilak, and C. Jacobs, "The H-index: use and overuse," J. Biomech. **44**, 208–209 (2011).

[15]G. Cormode, Q. Ma, S. Muthukrishnan, and B. Thompson, "Socializing the h-index," J. Informtr. **7**, 718–721 (2013).

[16]K. Dienes, "Completing h," J. Informtr. **9**, 385-397 (2015).

[17]L. Egghe, "Theory and practise of the g-index," Scientometrics. **69**, 131–152 (2006).

[18]P.R.C. Rahul, "r-index: Quantifying the quality of an individual's scientific research output," J. Scientometric Res., **2**, 80–82 (2013).

[19]R. Adler, J. Ewing, and P. Taylor, "Citation statistics," Stat. Sci. **24**, 1–14 (2009).

[20]M. Kosmulski, "Are you in h?," J. Informetrics. **7**, 693–698 (2013).

[21]M. Bordons, J. Aparicio, B. González-Albo, and A.A Díaz-Faes, "The relationship between the research performance of scientists and their position in co-authorship networks in three fields," J. Informetrics. **9**, 135–144 (2015).

[22]V. Bucheli, A. Díaz, J. Calderón, P. Lemoine, J. Valdivia, J. Villaveces, and R. Zarama, "Growth of scientific production in Colombian universities: an intellectual capital-based approach," Scientometrics. **91**, 369–382 (2012).

[23]B.S. Lancho-Barrantes, and F.J. Cantú-Ortiz, "Science in Mexico: a bibliometric analysis," Scientometrics. **118**, 499-517 (2019).

[24]S. Cole, and T.J. Phelan, "The scientific productivity of nations," Minerva, **37**, 1–23 (1999).

[25]A.M.C Şengör, "How scientometry is killing science," GSA Today, **24**, 44–45 (2014).

[26]J. Kim, H. Kim, and J. Diesner, "The impact of name ambiguity on properties of coauthorship networks," J. of infosci. Theory and Practice, **2**, 6–15 (2014).

[27]H. Wickham, *ggplot2: Elegant graphics for data analysis*, (Springer-Verlag, New York, 2016).

[28]R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing Vienna, Austria. (2017).

[29]M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An Open Source Software for Exploring and Manipulating Networks," Proceedings of the third International AAAI Conference on Web and Social Media, **3**, 361–362 (2009).

[30]G. Csardi, and T. Nepusz "The *igraph* software package for complex network research," InterJournal Complex Systems, 1695 (2006).

[31]A. Lotka, A., "The frequency distribution of scientific productivity," J. Washington Acad. Sci., **16**, 317–323 (1926).

[32]M. MacRoberts, and B. MacRoberts, "A Re-Evaluation of Lotka's Law of Scientific Productivity," Soc. Stud. Sci., **12**, 443–450 (1982).

[33]A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-Law Distributions in Empirical Data," SIAM Rev., **51**, 661–703 (2009).

[34]C. S. Gillespie, "Fitting Heavy Tailed Distributions: The *poweRlaw* Package," J. Stat. Softw, **64**, 1–716 (2015).

[35]M. Perc, "The Matthew effect in empirical data," J. R. Soc. Interface **11**, 20140378 (2014).

[36]M. van Steen, *Graph theory and complex networks: an introduction*, (van Steen, Lexington, 2010).

[37]C. Pozrikidis, *An introduction to grids, graphs, and networks* (Oxford University Press, Oxford, 2014).

[38]A. L. Barabási, *Network science* (Cambridge University Press, Cambridge, 2016).

[39]M. Newman, *Networks* (Oxford University Press, New York, 2018).

[40]E. Estrada, and P. Knight, *A first course in network theory* (Oxford University Press, Oxford, 2015).

[41]D. Magnolo, "Dynamical systems associated with adjacency matrices," Discrete Cont. Dyn-S., **23**, 1945–1973 (2018).

[42]G. Caldarelli, and M. Catanzaro, *Networks: a very short introduction* (Oxford University Press, Oxford, 2012).

[43]S. P. Borgatti, M. G. Everett, and J. C. Johnson, *Analyzing social networks* (Sage, Los Angeles, 2018).

[44]F. Menczer, S. Fortunato, and C. Davis, *A first course in network science* (Cambridge University Press, Cambridge, 2020).

[45]Z. Yang, R. Algesheimer, and C. Tessone, "A Comparative Analysis of Community Detection Algorithms on Artificial Networks," Sci. Rep. **6**, 30750 (2016).

[46]V. Subieta-Frías , and G. M. Ramírez-Ávila, "Scientometric study of the Revista Boliviana de Física with elements of complex networks analysis," Rev. Bol. Fis, **35**, 24–36 (2019).