Summer 5-8-2021

Analysis of citizen science scientific publications: A scientometric study

Ajay Kumar Chaubey Dr.
Government Lead P.G. College, Waidhan, Singrauli, M.P, ajaybhu21@gmail.com

Ankit Singh Mr.
DLIS, Banaras Hindu University, Varanasi, U.P, ankitsng2012@gmail.com

Analysis of citizen science scientific publications: A scientometric study

Dr. Ajay Kumar Chaubey
Librarian, Government Lead P.G. College, Waidhan, Singrauli, M.P
Orcid Id: https://orcid.org/0000-0003-4925-7723
Email: ajaybhu21@gmail.com

Ankit Singh
Junior Research Fellow,
Dept. Of Library and Information Science,
Banaras Hindu Universirty, Varanasi, U.P
Email: ankitsng2012@gmail.com

Abstract
The purpose of this study is to assess the citizen science publications in journals indexed in the Web of Science database from 1993 to 2020.Major areas of research related to CS are environmental science, ecology, biodiversity & conservation, education & educational research, public environmental & occupational science and computer science. The USA, England, Australia, Canada and Germany were the most productive countries. This study describes some silent characteristics trends of CS research like most prolific authors, institutions and countries. 82% of publications on CS received the citation while 18% of publications did not receive any citation. Although 47% of publications were open access and 53% publications were close access platform. However, Journal on Biological Conservation, Peerj, International Journal of Environment and Pollution, Marine Pollution Bulletin and Journal of Applied Ecology are the leading journals in this field.

Keywords: Citizen Science, Scientometrics study, Vosviewer, Social Network analysis.

1.0 Introduction
Citizen science was introduced in the 1990s by Alan Irwin (Irwin, 1995) who mentioned that science should fulfil the need to empower people and society. Oxford English Dictionary (OED) adopted the term citizen science in June 2014.In general, citizen science concern with the active participation and collaboration of common people in scientific research works to produce new knowledge for society. Citizen science engaged researcher working with non-professional or amateur scientist. It is also called public participation research. Now citizen science has been well known in the field of natural science and local history. The emergence of the internet has immensely expanded the opportunity for the civic to involve in scientific research using CS research. These days tremendous growth has been observed in the number of institutions, government agencies and research funders, taking an interest in the field of citizen science. This interest is a positive sign for citizen science research. This paper aims analysing the evolution and collaboration networks of citizen science publications published in Web of Science from 1993 to 2020.Therefore, the growth and extension of citizen science are here characterized quantitatively employing the study of citation characteristics, co-occurrence of keyword and co-authorship among scientists, Institutions and countries.

2.0 Review of Literature
Scientometrics becomes very perspective research for measuring the national as well as international research outcome for uncovering the network of authors, institutions, journals and future research priorities etc. Bonney et al. (2009) have carried out a study of the activities undertaken in citizen science projects, mapping their academic impact through a set of quantitative indicators and found that citizen science project is increasing in number day to day. Kumar (2015) has mentioned in his study that network Visualization forms a significant component of network analysis. It provides meaning to the analysis. Co-authorship networks

at the institutional and international level have rapidly grown during the last decade and also observed that in association with other indicators of scientometrics like citation, co-citation, co-word analysis acquaintance, assortative mixing patterns including various socio-academic parameters, among others. Follet and Strezov (2015) studied the SCOPUS and Web of Science databases to identify and analyse publications on citizen science and their application in new research projects. Kullenberg & Kasperowski (2016) in their study mentioned that the largest scientific publications identified in the field of ornithology, astrology, meteorology and microbiology. CS research found in the field of biology, ecology and conservation. However, Turrini et al. (2018) pointed that in citizen science research public participation was considerably less important in terms of creating new knowledge and learning opportunities. Bautista-Puig et al. (2019) have conducted a study on the scientific landscape of citizen science and find out that open access documents 30.7 % higher than other documents and citizen science area (Health, Bio, Geo and Public) have been rapidly grown after 2010.In another literature review, Pelacho et al. (2021) undertook the Analysis of the evolution and collaboration networks of citizen science scientific publications. This study revealed a tremendous growth in the number of publications per year. A large number of researchers consider citizen science to be an appropriate methodology in their area of interest.

3.0 Objectives

The objective of the present study is to explore the research trend on CS research in terms of (i) growth of literature (ii) citation characteristics of CS research (iii) prominent authors, institutions and countries (iv) productivity of most relevant journals.

4.0 Method of Study

To assess the global research trend in citizen science (CS) research, we conducted scientometric and social network analysis methods. The bibliometrics data was retrieved from the Web of Science (WoS) Core Collection [1993-2000].VOSviewer (version 1.6.16) was applied for network visualization. Web of Science is one of the most prominent and extensive database. It contains high-quality publications that can be analysed and visualized for bibliometrics study.

Search strategy

Web of Science core collection basic search keyword used Citizen Science OR community science OR crowd science OR crowd-sourced science OR civic science OR volunteer monitoring OR Participatory science) in the topic search. Although the bibliometric parameters of the Web of Science output (2872 publications) have been downloaded in tab-delimitated files and then imported in MS Excel file for analysis.

5.0 Results and Discussion

5.1 Year-wise growth of publications

Figure 1 depicts the CS publications that were published from 1993 to 2020. Although we have considered publications that appeared from 1993 to 2020, to get a vivid picture we have considered citation that the publications received up to the end of 2020. It is clearly shown in the figure, CS literature increased rapidly from 2015 to 2020. It is also observed that there are no publications in the year 1996. It is a notable growth which is enough to prove that CS has obtained much attention from scientists, researchers and practitioners from around the globe.

5.2 Subject Specialization

As indicated in table 1 and figure 2, we categorised the top 15 research area and their number of publications across the period studied (1993-2020) which were Environmental Sciences & Ecology with 667 documents, Biodiversity & Conservation with 411 documents, Education & Educational Research and computer science with 123 documents, Health Care Sciences & Services with 55 documents, Communication and Life Sciences & Biomedicine with 53 documents and Psychology with 50 documents.

Figure 1: Year wise growth of publications

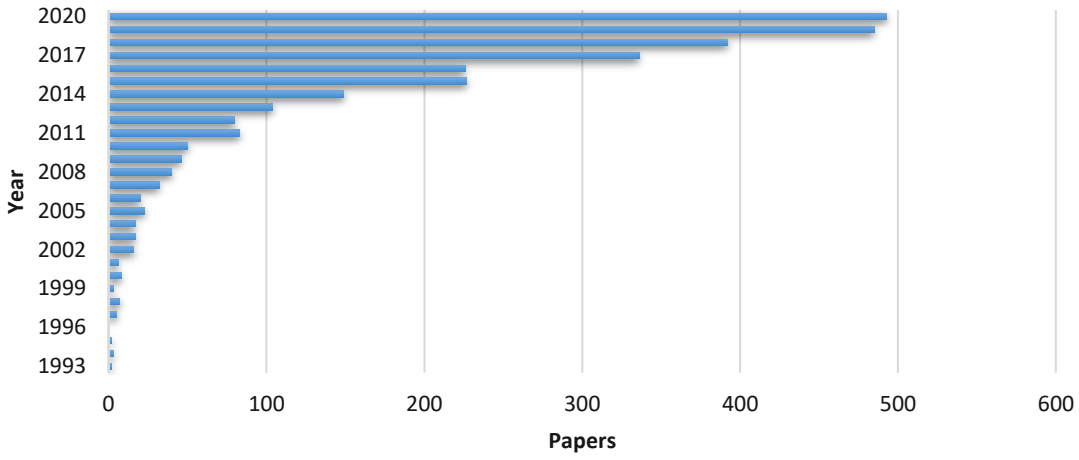


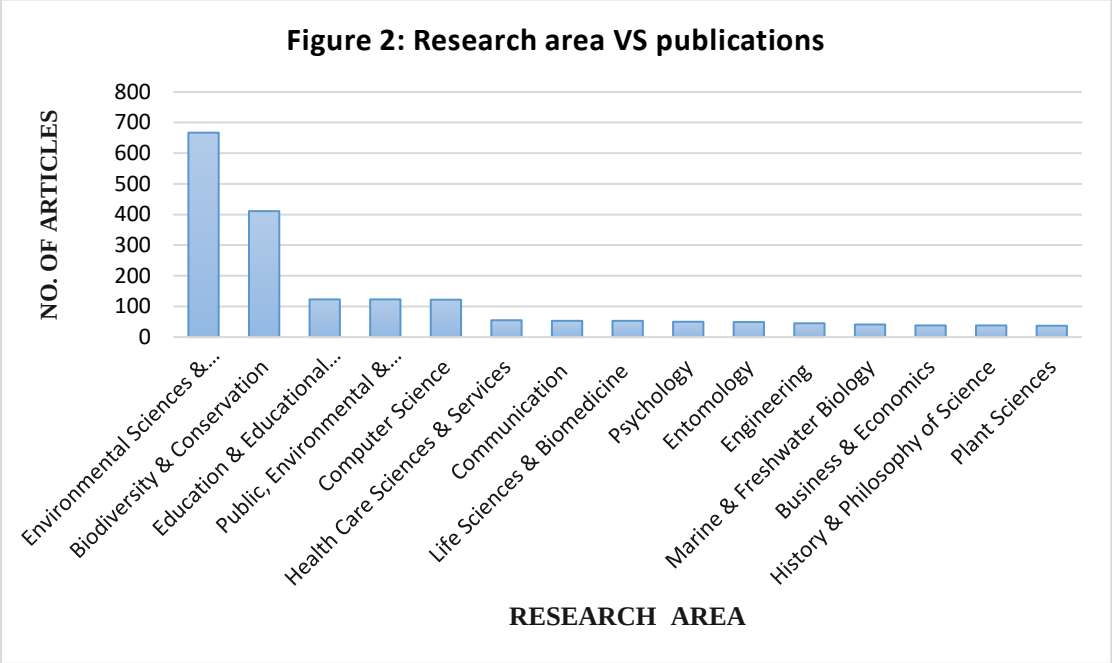
Table1: Top-15 Research areas Vs number of publications

Research Area	Publications
Environmental Sciences & Ecology	667
Biodiversity & Conservation	411
Education & Educational Research	123
Public, Environmental & Occupational Health	123
Computer Science	122
Health Care Sciences & Services	55
Communication	53
Life Sciences & Biomedicine	53
Psychology	50
Entomology	49
Engineering	45
Marine & Freshwater Biology	41
Business & Economics	38
History & Philosophy of Science	38
Plant Sciences	37

5.3 Citations Characteristics of CS

A citation appearing in an article is an indication of the information usage**.** However, exiting citation studies are based on WOS database. Table 2 & figure 3 analysed the cited vs non-cited publications from 1993 to 2020. Of the total 2872 publications received 42954 citations till December 2020. It was found that 82.07% of papers received citations whereas 17.93% of the total publications did not receive any citations till December 2020. In 2015 CS

publications were received the highest number of citations (5437) followed by 2017 (4399 citations).



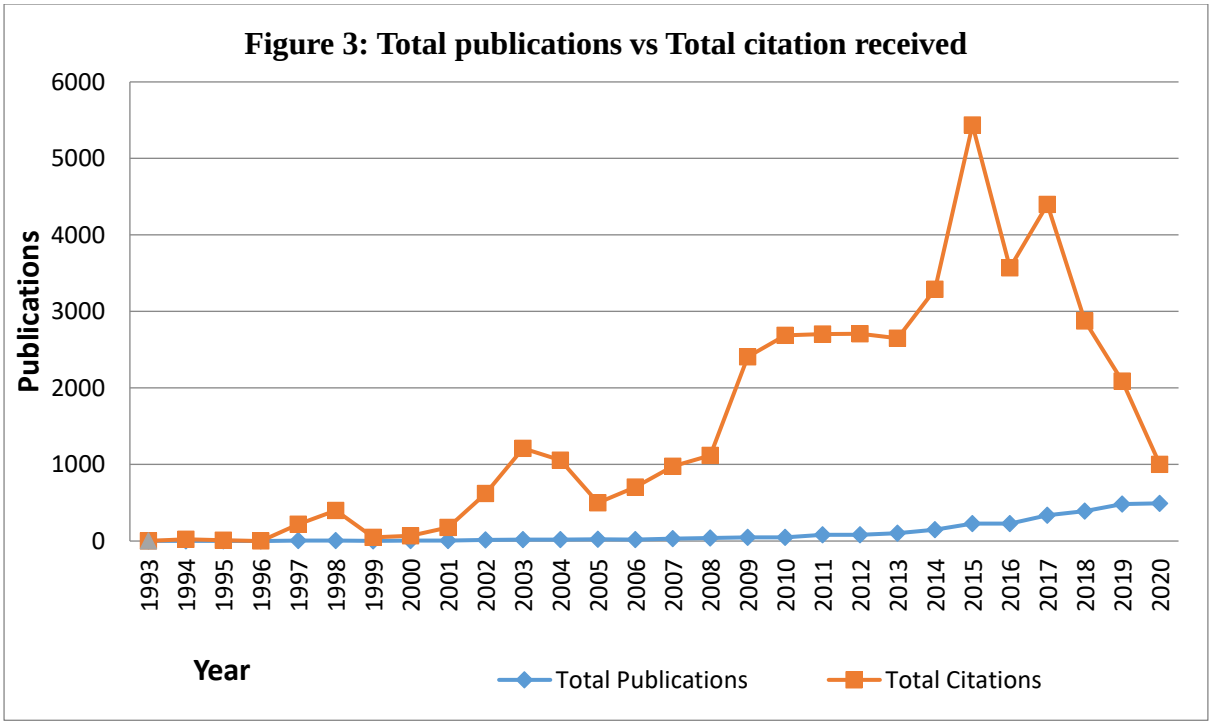


Table 2: Cited VS Non-cited publications on CS

Year	TP	TC	Total no. of the cited Pub.	% of the Cited pub.	Total no. of non-cited Pub.	% of thenon-citedPub.
1993	2	1	1	50.00	1	50.00
1994	3	25	2	66.67	1	33.33
1995	2	11	1	50.00	1	50.00
1996	0	0	0	0	0	0
1997	5	219	3	60.00	2	40.00
1998	7	400	5	71.43	2	28.57
1999	3	46	2	66.67	1	33.33
2000	8	69	3	37.50	5	62.50
2001	6	176	6	100.00	0	0.00
2002	16	620	10	62.50	6	37.50
2003	17	1212	12	70.59	5	29.41
2004	17	1055	14	82.35	3	17.65
2005	23	499	15	65.22	8	34.78
2006	20	707	14	70.00	6	30.00
2007	32	976	26	81.25	6	18.75
2008	40	1121	28	70.00	12	30.00
2009	46	2406	35	76.09	11	23.91
2010	50	2690	46	92.00	04	8.00
2011	83	2705	70	84.34	13	15.66
2012	80	2707	72	90.00	08	10.00
2013	104	2649	90	86.54	14	13.46
2014	149	3289	127	85.23	22	14.77
2015	227	5437	191	84.14	36	15.86
2016	226	3570	196	86.73	30	13.27
2017	336	4399	279	83.04	57	16.96
2018	392	2876	321	81.89	71	18.11
2019	485	2088	413	85.15	72	14.85
2020	493	1001	375	76.06	118	23.94
1993-2020	**2872**	**42954**	**2357**	**82.07**	**515**	**17.93**

TP= Total Publications, TC=Total Citations, no. =number, pub. =publications

Figure 4: Share of publications output on CS by publication year: 1993–2020

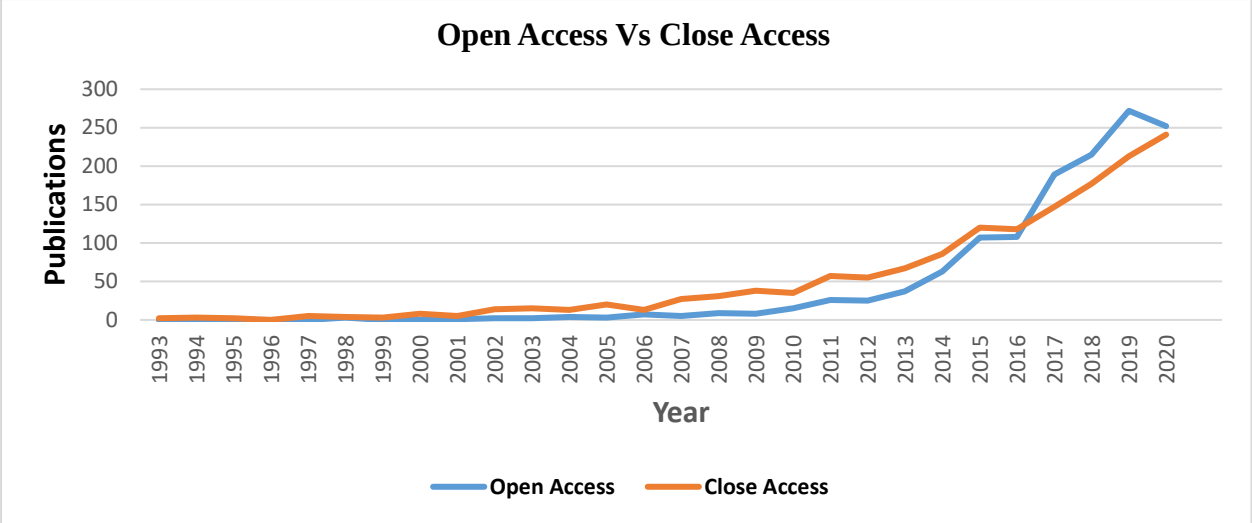


Table 3: Open access Vs Close access publication on CS

Year	Total Publication	Open Access	% of the Open Access	Close Access	% of the Close Access
1993	2	0	0.00	2	100.00
1994	3	0	0.00	3	100.00
1995	2	0	0.00	2	100.00
1996	0	0	0%	0	0%
1997	5	0	0.00	5	100.00
1998	7	3	42.86	4	57.14
1999	3	0	0.00	3	100.00
2000	8	0	0.00	8	100.00
2001	6	1	16.67	5	83.33
2002	16	2	12.50	14	87.50
2003	17	2	11.76	15	88.24
2004	17	4	23.53	13	76.47
2005	23	3	13.04	20	86.96
2006	20	7	35.00	13	65.00
2007	32	5	15.63	27	84.38
2008	40	9	22.50	31	77.50
2009	46	8	17.39	38	82.61
2010	50	15	30.00	35	70.00
2011	83	26	31.33	57	68.67
2012	80	25	31.25	55	68.75
2013	104	37	35.58	67	64.42
2014	149	63	42.28	86	57.72
2015	227	107	47.14	120	52.86
2016	226	108	47.79	118	52.21
2017	336	189	56.25	147	43.75
2018	392	215	54.85	177	45.15
2019	485	272	56.08	213	43.92
2020	493	252	51.12	241	48.88
1993-2020	**2872**	**1353**	**47.11**	**1519**	**52.89**

5.4 Availability of CS publications

In the next step, attempts were made to understand the pattern of CS publications in terms of availability of content. As we can see in table 3 & figure 4 majority of publications were close access. Of the total, 52.89% were close access while 47.11% were open access in CS. The interesting point here to note that in the year 2017, 2018, 2019 & 2020 open access publications on CS were more than close access. The reason behind the increase of open access publications on CS is in the last few years open access becomes more unrestricted, more data is becoming available to explore the impacts.

6.0 Social Network analysis

Social network analysis (SNA) in general is mapping and measuring the nature of the individual at the micro-level, the pattern of relationships (network visualization) at the macro level, and the relations between the two. Network visualization can reveal the state of affair and development status of the discipline.SNA also provides both a visual and mathematical analysis of groups, institutions, peoples and other connected information entities. We use network visualization through VOSviewer (www.vosviewer.com).In the VOSviewer map, each node shows a term and the size of the node and font represents the activity of the term. The larger the node and font, the more active the term is in the field, and vice versa. The distance between any two nodes in the figures exhibits the degree of association between the two terms. The smaller the distance between the nodes of the two terms, the stronger the correlation between the two terms, and vice versa. Each point in a diagram has a colour that depends on the density of items at that point and also many colours on the map, represents the diversification of research direction. A similar colour shows a close relationship between two terms and the nodes with the same colour belong to a cluster.

6.1Subject Analysis by Co-Occurrence of author Keywords

The number of times that the author keyword appeared and their dynamics throughout the period were analysed. The raw keywords have been cleaned up. The bibliographic data represent that there are 10005 author keywords obtain in the publications. The co-occurrence threshold of keywords was set to 10 which led to getting 160 keywords in VOSviewer. As indicated in figure 5, all keywords are grouped into 10 clusters. Cluster 1 is represented by red colour that primarily deals with concepts like ‘education’ (29 links, 53 link strength), community (28 links, 61 link strength), science(28 links, 58 link strength), science communication (27 links, 46 link strength), sustainability (27 links, 42 link strength) and public participation (25 links, 60 link strength). Cluster 2 is depicted by green colour that deals with the concepts like citizen science (140 links, 1882 link strength), conservation (54 links, 153 link strength) and ebird (27 links, 76 link strength).

Table 4: The top ten high-frequency author keywords in CS

Keyword	Occurrence	Total Link Strength
Citizen Science	1832	1882
Crowdsourcing	110	182
Monitoring	108	207
Implementation Science	96	69
Climate Change	92	162
Conservation	72	153
Community	58	61
Biodiversity	57	133
Community Engagement	56	95
Community-based Participatory Research	55	92

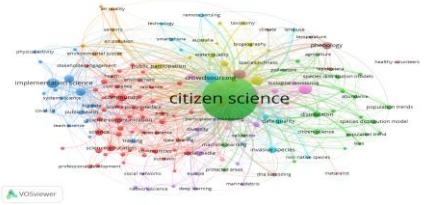


Figure 5: Network visualization map of high-frequency Author Keywords

Cluster 3 is indicated by blue colour dealing with concepts like community engagement (43 links, 95 link strength), Community base participation (36 links, 92 link strength) and implementation science (22 links, 69 link strength).Cluster 4 is indicated by yellow colour represents concepts like monitoring (63 links, 207 link strength), biodiversity (55 links, 133 link strength) and volunteer (31 links, 75 link strength). Cluster 5 is represented by purple colour which represents concepts like biodiversity monitoring (34 links, 63 link strength) and diversity (21 links, 31 link strength) etc.

6.2 Most active author, country and institutions on CS research

In this section we have analysed the co-authorship network visualization of CS literature in terms of authors, country and institutions. Detailed analysis is given below.

6.2.1 The Co-Authorship Analysis of CS articles by authors

Co-authorship is one of the most distinct forms of modern research collaboration. A co-authorship network is a social network in which two or more scientists are engaged in one or more publications through an indirect path linked to each other. The co-authorship network of CS is depicted in figure 6. While the minimum number of papers published by an author was set at 5, 89 authors with a 456 total link strength were figured. All 89 authors were divided into 8 clusters. As indicated in table 5 & Figure 6 maximum collaboration was observed among Fink, D. From Cornell University, Hochachka, W.M. from Cornell University, Callaghan, C.T. from University of New South Wales, Roy, D.B. from UK Centre for Ecology & Hydrology, King, A.C. from Stanford University and Kelling, S. from Cornell University. However, Fink, D and Hochachka, W.M. have written (17 articles and 15 articles, respectively) with 16 co-authors. Callaghan, C.T has written 15 articles with 6 co-authors and Roy, D.B. has written 14 articles with 13 co-authors. These authors showed strong collaboration and influential work in the CS field.

Table 5: Top ten most prolific authors in CS research

Authors	Affiliation	Doc	Cit	LS
Fink, Daniel	Cornell University, USA	17	1115	52
Hochachka,Wesley,M.	Cornell University, USA	15	645	51
Callaghan, Corey,T.	University of New South Wales, Australia	15	100	26
Roy, David, B.	UK Centre for Ecology & Hydrology, UK	14	546	22
King, Abby, C.	Stanford University,USA	13	81	07
Kelling , Steve	Cornell University, USA	12	1811	50
Zuckerberg, Benjamin	University of Wisconsin–Madison, USA	12	401	14
Porfiri, Maurizio	New York University, USA	10	117	11
Gillings, Simons	British Trust for Ornithology, UK	10	92	10
Bonney, Rick	Cornell University, USA	09	1787	33

Doc =Document, Cit=Citations, LS=Link Strength

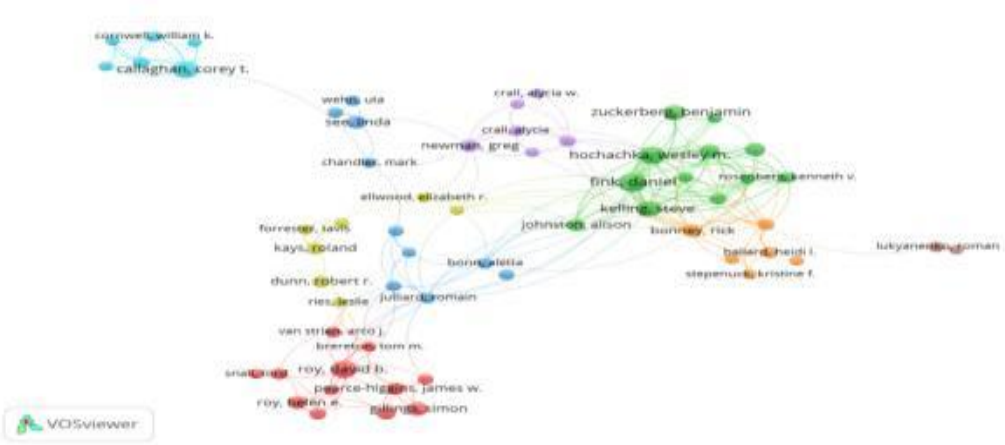


Figure 6: Co-authorship network of the most productive authors in CS

6.2.2 The Co-Authorship analysis of CS articles by Institutions

Every institution has researcher or scientist ("nodes") who serve as a critical channel for the exchange of ideas and information. SNA reveals the institutional interactions within and across the country and it also identifies the similar patterns of connections among institutions. The institutions' visualization map is depicted in Figure 7. On setting a minimum threshold of 5 publications of an institution, out of the total 3821 institutions, 355 institutions were under the threshold. University of Washington, Cornell University, University of Florida, University of Wisconsin, US Geological Survey and the University of Queensland are the topmost influential institutes of CS research in the world in terms of the total number of publications, University of Washington contributed 60 publications and co-authorship with 85 institutions, Cornell University contributed 52 publications and co-authorship with 63 institutions, University of Florida contributed 48 publications and co-authorship with 72 institutions, University of Wisconsin contributed 44 publications and co-authorship with 45 institutions and US Geological Survey contributed 43 publications and co-authorship with 60 institutions.

Table 6: Top ten institutions publishing on CS

Institutions	Document	Citations	Link strength
University of Washington	60	1247	129
Cornell University	52	2195	84
University of Florida	48	519	101
University of Wisconsin	44	1242	72
US Geological Survey	43	1330	87
University of Queensland	42	885	116
University of Oxford	41	1201	85
University of Minnesota	37	804	87
Colorado State University	35	1051	76
University of California Devis	35	1255	67

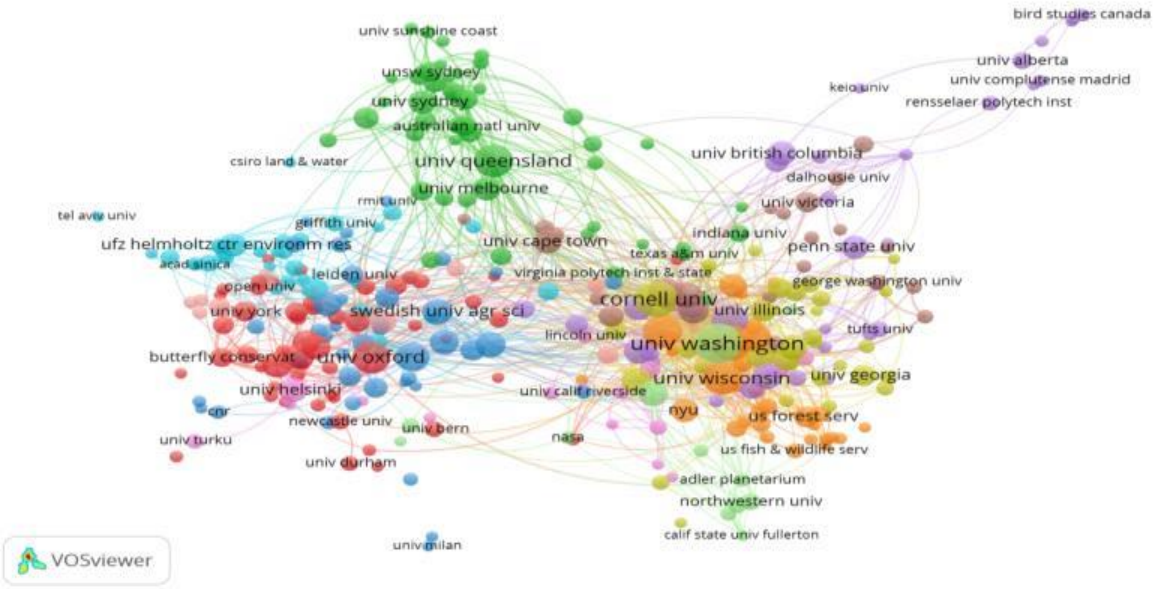


Figure 7: Co-authorship network between top institutions in CS

6.2.3 Co-Authorship Analysis of Countries

Country co-authorship analysis is an important form of co-authorship analysis. It can reveal the degree of communication between countries as well as the influential countries in this field. The threshold for minimum number of papers published by a country was set at 25. Of the total 121 countries, 31 countries with a 2203 total link strength were figured and divided into 4 clusters. All 31 countries were reflecting solid international collaborative research. As we can see in table 7 and figure 8, the most influential countries, according to their degree centrality, were the USA, England, Australia, Canada and Germany, as represented by the larger nodes. The USA contributed 1283 publications with 23864 citations, England contributed 456 publications with 8325 citations, Australia contributed 275 publications with 3518 citations and Canada contributed 233 publications with 4406 citations.

Table 7: Top ten most productive countries in CS

Country	Document	Citations	Link strength
USA	1283	23864	675
England	456	8325	572
Australia	275	3518	282
Canada	233	4406	230
Germany	185	2766	370
France	149	2946	268
Netherland	138	2283	271
Spain	125	1029	195
Italy	125	1425	165
Sweden	85	1274	152

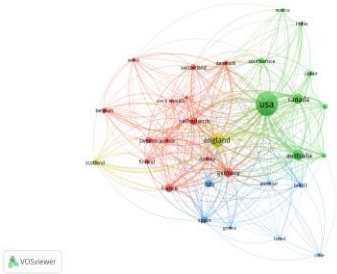


Figure 8: Co-authorship network between countries in CS

6.3 Bibliographic Coupling

The concept of Bibliographic coupling was first given by Kessler in 1963. It is used to track and trace similar pattern between pairs of papers, authors and organisations. For network visualization we used VOSviewer. With the help of VOSviewer, we can trace the journals which are most strongly connected (focal) to CS research and identify the related journals of this field. On setting a minimum threshold of 5 publications of a journal of the total 1044 journal, 132 journals were selected for visualization on the map. The top ten journals, their citation, articles and link strength are shown in table 8.As we can see in the table 8 and figure 9, journal on Biological Conservation, Peerj, International Journal of Environment and Pollution, Marine Pollution Bulletin, Journal of Applied Ecology, Biodiversity and Conservation, Ecology and Society, Science of the Total Environment, Ecosphere, and Sustainability are the leading sources of publication on the CS research. These journals also obtained the highest link strength. It means that these journals represent the main channel for further publications in this field. It is

found that other high influential journals like Ecological Applications (21 articles with 7450 link strength), ISPRS International Journal of Geo-Information (21 articles with 6286 link strength), Conservation Biology (20 articles with 10620 link strength) and Journal of Environmental Management (20 articles with 10031 link strength) also published research publications in this field.

Table 8: Top ten Journals according to publications

Journal Name	Art. (n=2872)	% of Art.	Citation (n=42954)	% of Citation	TLS
Biological Conservation (IF=4.71)	71	2.47	2996	6.97	40487
Peerj (IF=2.38)	40	1.39	324	0.75	9352
International Journal of Environment and Pollution (IF=0.54)	33	1.15	169	0.39	3023
Marine Pollution Bulletin (IF=4.04)	32	1.11	545	1.27	9517
Journal of Applied Ecology (IF=5.84)	30	1.04	814	1.90	12361
Biodiversity and Conservation(IF=2.93)	28	0.97	729	1.70	12135
Ecology and Society (IF=4.14)	26	0.91	1511	3.52	10175
Science of the Total Environment(IF=6.55)	26	0.91	336	0.78	7259
Ecosphere (IF=2.87)	26	0.91	197	0.46	9204
Sustainability(IF=2.57)	22	0.77	50	0.12	9086

Note: Impact factor (IF) is based on JCR IF, 2019.Art. =Articles, TLS=Total link strength

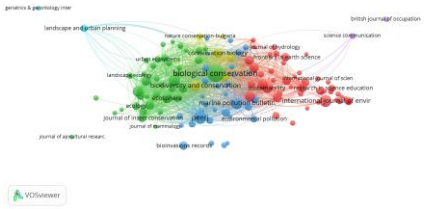


Figure 9: Bibliographic Coupling network of CS Journals

Conclusion

The present study revealed that there is a rapid expansion in the research activities related to CS after 2015.An analysis of 2872 publications from the WoS database was done, and as a result, the top ten keywords, research area, journals and most prolific authors, institutes and

countries were identified. As indicated in the analysis, the USA, England, Australia, Canada and Germany were the most active countries in the CS field. Fink, Daniel from Cornell University, Hochachka, Wesley, M. from Cornell University and Callaghan, Corey,T. from The University of New South Wales had the most articles. Citizen science, crowdsourcing, monitoring, implementation science, climate change and conservation were high-frequency keywords in CS. However, Biological Conservation, Peerj, International Journal of Environment & Pollution and Marine Pollution Bulletin were the leading journals in this field. Although the major area of a study employing CS is to be found in Environmental Sciences & Ecology, Biodiversity & Conservation, Education & Educational Research Public, Environmental & Occupational Health and Computer Science.
Overall, this study provides helpful insight for the researcher and practitioner in this domain and revealed the global research trends of CS research.

References

1) Bautista-Puig, N., De Filippo, D., Mauleón, E., & Sanz-Casado, E. (2019). Scientific landscape of citizen science publications: Dynamics, content and presence in social media. *Publications*, 7(1), 12.
2) Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience*, 59(11), 977-984.
3) Follett, R., & Strezov, V. (2015). An analysis of citizen science based research: usage and publication patterns. *PloS one*, 10(11), e0143687.
4) Irwin, A. (1995). Citizen science: A study of people, expertise and sustainable development.Environment and society. London: Routledge.
5) Kumar, S. (2015). Co-authorship networks: A review of the literature. *Aslib Journal of Information Management*, 67(1), 55–73.
6) Kullenberg, C., & Kasperowski, D. (2016). What is citizen science?–A scientometric meta-analysis. *PloS one*, 11(1), e0147152.
7) Morel, C. M., Serruya, S. J., Penna, G. O., & Guimarães, R. (2009). Co-authorship network analysis: a powerful tool for strategic planning of research, development and capacity building programs on neglected diseases. *PLoS Negl Trop Dis*, 3(8), e501.
8) Pelacho, M., Ruiz, G., Sanz, F., Tarancón, A., & Clemente-Gallardo, J. (2021). Analysis of the evolution and collaboration networks of citizen science scientific publications. *Scientometrics*, 126(1), 225-257**.**
9) Turrini, T., Dörler, D., Richter, A., Heigl, F., & Bonn, A. (2018). The threefold potential of environmental citizen science-Generating knowledge, creating learning opportunities and enabling civic participation. *Biological Conservation*, 225, 176-186.