MJFAS
Malaysian Journal of
Fundamental and Applied
Sciences

RESEARCH ARTICLE

Understanding Published Literatures on Persistent Homology using Social Network Analysis

Muhamad Luqman Sapini[a,b,*], Mohd Salmi Md Noorani[a], Fatimah Abdul Razak[a], Mohd Almie Alias[a], Norliza Muhamad Yusof[b]

[a] Department of Mathematical Sciences, Faculty of Sciences and Technology, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia; [b] Universiti Teknologi MARA (UiTM), Cawangan Negeri Sembilan, Kampus Seremban, Seremban 70300 Negeri Sembilan, Malaysia

Abstract In several fields, topology is well adapted for analyzing big data and much more potent than conventional data analysis methods. Persistent Homology is an algebraic method used in topological data analysis for measuring shapes and functions. This paper aimed to analyze literature on Persistent Homology. From its inception (2004) to 6th November 2020, we comprehensively searched the Scopus science database. This paper records the document and source types, year of publication, languages of documents, subject areas, author's keywords, country/region, authors, institutions, and citations, utilizing standard bibliometric indicators. The VOSviewer applications were used for analysis through various network forms. Two network measurements were considered to analyze the network model, which are degree centrality and betweenness centrality. Additional analysis was undertaken using Persistent Homology itself to comprehend the co-authorship network better. The study found that the number of Persistent Homology research has risen dramatically since 2014. Globally, scholars in the area mainly come from the United States. Regarding the number of citations, with an average of 51.2 citations each year, the paper by Zomorodian and Carlsson (2005) appears as the most cited article. Much of the study on Persistent Homology was in the fields of computer science and mathematics. The most used keywords were "topological features" and "data handling", which reflect the significant study fields covered in Persistent Homology. Overall, the increasing number of Persistent Homology publications reflects a growing knowledge of its meaning and unique demands.

Keywords: persistent homology, topological data analysis, bibliometric analysis, social network analysis, centrality.

***For correspondence:**
luqman0211@uitm.edu.my

Received: 13 Sept 2021
Accepted: 25 August 2022

Introduction

The rising number of available data has resulted in the advancement of knowledge processing methods other than machine learning approaches. Big data provides us with the source to extract new insights in various disciplines [1]. Topological data analysis (TDA) is a recent field [2] that arose during the first decade of the century from various works in applied (algebraic) topology and computational geometry. TDA is a novel statistical technique, particularly powerful for the analysis of big data and high dimensional data sets [3]. The underlying TDA principle is that shape matters.

Persistent Homology is a tool used in TDA to analyze qualitative data characteristics that persist in multiple scales [4]. Persistent Homology is an algebraic method for measuring shapes and functions with topological features [5]. Homology has to do with measuring such topological properties, while persistence refers to evaluating which of these are retained in the data over multiple scales [6]. Persistent

Homology offers a basis for defining a static data set's topology [7]. In several fields, topology is well adapted for analyzing big data and much more powerful than conventional methods of data analysis [8].

Persistent Homology began with [9] ground-breaking works as a method of TDA. Persistent Homology will bridge the gap between topology and geometry and open-up new possibilities for mathematics, computer science, computational biology, biomathematics, and engineering researchers [10]. The simplicial complex method is the basic idea of Persistent Homology so that each topological generator is fitted with a geometric measurement. Ref. [4] summarizes the pipeline for the computation of Persistent Homology from data in Figure 1.

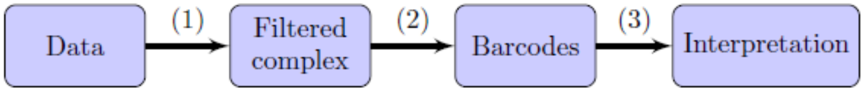

Figure 1. Persistent homology pipelines.

There have been number of research to explore the current state of Persistent Homology. For the past five years, Persistent Homology has been used in various fields such as brain signal [11]–[18], social and complex network [7], [19]–[21], shape analysis [22]–[25], biological data analysis [26]–[28], image processing [29]–[32], as well as other fields. The use of Persistent Homology in time series applications is still new and growing rapidly. Ref. [33] presents a new chaos detection method based on Persistent Homology, Ref. [34]–[36] explored the use of Persistent Homology on haze time series while [40]-[41] used Persistent Homology on financial time series. Ref. [39] proposes a hybrid approach integrating methods of machine learning with persistent Homology to improve prediction efficiency.

[9] initiated the study of Persistent Homology in 2004. Since then, researchers' development of the topological tool in the related industry has been widely used by researchers to further investigate Persistent Homology. With growing numbers of Persistent Homology studies, it is useful to observe those studies' general patterns. Therefore, the objective of this study is to examine the scientific literature published using bibliometric analysis in the field of Persistent Homology.

Bibliometric analysis is a quantitative approach based on the analysis of relevant publications to analyze the research fields' knowledge structure and growth [40]. [41] defines bibliometrics as "the application to books and other media of communication of mathematics and statistical methods" [42].

The methodology adopted is detailed in the second section. The third section describes the outcomes of the bibliometric variables concerned. The last section summarizes the analysis, discusses potential research areas, and presents certain limitations of this report.

Methods

Data Sources and Search Criteria Process

In this research, scientific studies published in the Persistent Homology area and included in the Scopus database were analyzed using the bibliometric analysis method. We utilized a Scopus database and conducted a topic search as of 6th November 2020.

We focus on all Persistent Homology documents based on the title and keyword of papers for this study. The following query was performed: (TITLE ("persistent homology" OR "persistence homology") OR KEY ("persistent homology" OR "persistence homology")).

We generated a total of 760 documents via this search. However, the data obtained have several duplicates. By excluding these duplicates, 756 scientific records remained for further data analysis. Any observations and findings such as publication year, author, subject area, keywords, affiliation, country/region, and language were extracted from the documents received.

Data Analysis

In this research, data processing and analysis were implemented by using Microsoft Excel 365 and Harzing's Publish or Perish software. Data that has been obtained from Scopus will be aggregated and analyzed using Microsoft Excel 365 through related figures and tables. Harzing's Publish or Perish software [43] is used to retrieve and evaluate citations from Scopus and Google Scholar.

The critical approach used to evaluate co-occurrence is Social Network Analysis (SNA). SNA is the method of exploring social systems through the use of networks. In this regard, VOSviewer software was used to assess the co-cited authors, regions, and keywords to visualize social networks [44]. In order to create the network, fractional counting and association strength will be used respectively as the counting method and normalized method in VOSviewer settings.

Centrality metrics are one of the most common methods for characterizing network nodes [45]. Centrality is such a significant index as it shows which node in one whole network takes a crucial role. Among the tools in the Centrality Metrics are Degree Centrality (DC), Closeness Centrality (CC) and Betweenness Centrality (BC) [46]. In this research, DC and BC are used to analyze the network model. DC of the node in a network is the total number of edges connected to it [47]. It can define as

$$DC(v) = \deg(v) \tag{1}$$

where deg(*v*) is the number of the link or edge that connects node *v*.

BC of the node in a network is the shortest path (geodesics) available between a pair of nodes. If the vertex or node locates in the only direction that other nodes must go through, then this node should be essential and has a strong betweenness centrality. The higher the value of BC of the node in a network, the more important a node for the information flow [45]. BC can define as

$$BC(v) = \sum_{j \neq k} \frac{g_{jk}(v)}{g_{jk}} \tag{2}$$

$g_{jk}(v)$ is the number of shortest routes connecting the nodes *j* and *k* with the nodes *v* and g_{jk} is a whole number of shortest paths connecting nodes *j* and *k*.

We performed an additional analysis using Persistent Homology to better understand the co-authorship network. Persistent Homology is used to measure the shape of the resulting network based on topological features. For more elaborate details on the persistent homology of a weighted network, refer to [7] and [48]. Clique complex and persistent diagram will be used respectively as filtered complex and interpretation in the Persistent Homology. For the persistence computation, we used Scikit-TDA [49] run in PYTHON software.

Analysis and Findings

Data collected were analyzed to identify document and source types, year of publication, languages of documents, subject areas, author's keywords, country/region, authors, institutions, and citations. We represented the findings as frequency and percentage.

Document and Source Types

This research found nine types of Persistent Homology documents: article, conference paper, book chapter, review, note, book, erratum, letter, and undefined type. Table I indicates that the proportion of articles published was 53.84 percent of documents, followed by 41.80 percent of conference papers. The categories of records in the book chapter collectively accounted for about 3%. Review, note, book, erratum, letter, and undefined were the lowest six categories, with less than one percent each.

Table I. Document types

Document types	Frequency	Percentage (%)
Article	407	53.84
Conference paper	316	41.80
Book Chapter	21	2.78
Review	6	0.79
Note	2	0.26
Book	1	0.13
Erratum	1	0.13
Letter	1	0.13
Undefined	1	0.13
Total	756	100.00

Table II. Source types

Source types	Frequency	Percentage (%)
Journal	429	56.75
Conference proceeding	228	30.16
Book	92	12.17
Book Series	7	0.93
Total	756	100.00

Table II presents four types of sources relevant to Persistent Homology that have been issued. Journals (56.75 percent) are the highest represented category, followed by conference proceedings with 30.16 percent. A significant number of documents, 12.17%, refer to books. With less than one percent, the lowest contribution came from the book series.

Year of Publications

The first article on Persistent Homology was published in 2004 by [9] titled "Computing the Persistent Homology." Relevant publications' development then steadily accelerated from 2008 to 2013. It is increased significantly since. Based on Figure 2, the number of papers slightly decrease in 2015 and 2018 compared to 2014 and 2017, respectively. For the year 2019, the number of publications produced is the highest, with an increasing percentage of close to 30 percent compared to the year 2018. This number remains the same as of 6th November 2020. Although the year 2021 has not yet arrived, the Scopus database has already published one article.

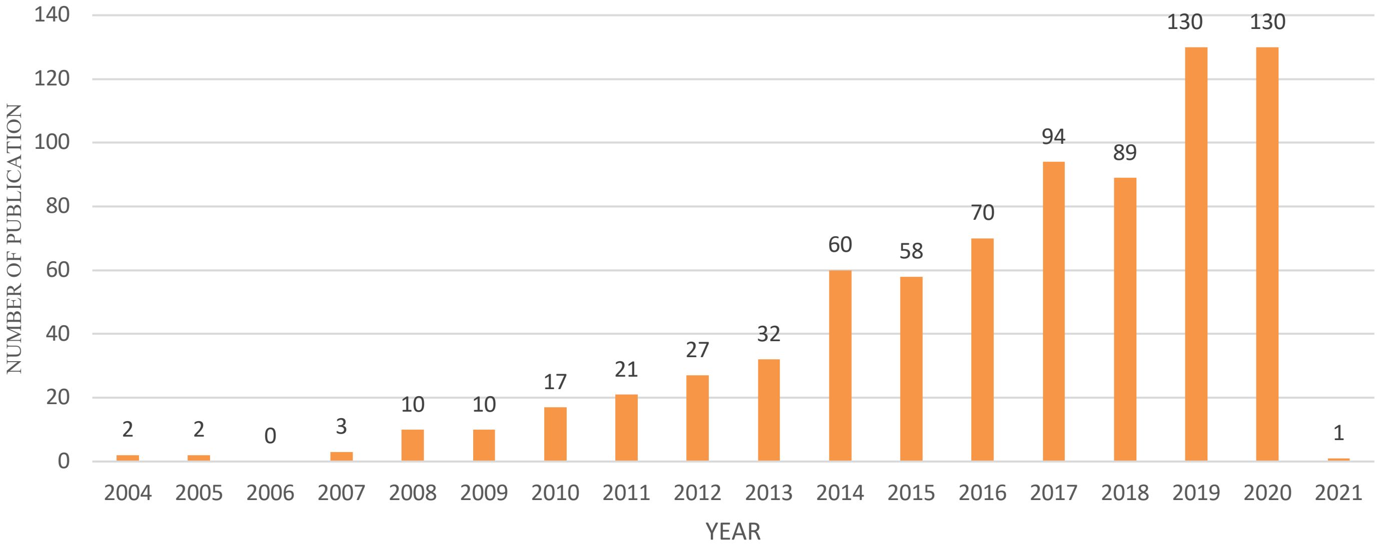

Figure 2. Total number of publications by year

Language of Documents

Table III indicates that most (more than 90 percent) of the acquired documents were written in English (99.34 percent). Any other publications were written in a single language of less than 1 percent, including Russian, Malay, and Chinese.

Table III. Language of Documents.

Language	Frequency	Percentage (%)
English	751	99.34
Russian	3	0.40
Malay	1	0.13
Chinese	1	0.13
Total	756	100.00

Subject Area

This research also presented documents based on their subject areas. Table IV tabulates the subject areas covered in the Persistent Homology. Since Persistent Homology is more focused on studies related to computer science and mathematics, it can be found that both subject areas were representing 60.88% and 53.40% of the total publications, respectively. It is followed by Engineering (20.29%), Physics and Astronomy (10.01%), and Biochemistry, Genetics and Molecular Biology (5.74%). This shows the variety of subjects where Persistent Homology can be applied on.

Keyword Analysis

The author keywords were mapped and visualized using VOSviewer software, a bibliometric network construction, and visualization software tool. We excluded the “Persistent Homology” and “Topological Data Analysis” keywords since they were the keywords used in the query. Figure 3 displays a network view of the author’s keywords in which nodes represent keywords, edges thickness represents relationships with other keywords (co-occurrence) and colour represent the cluster. Keywords of the same colour, for example, were in the same cluster. Thus, machine learning, learning systems, and artificial intelligence have a common colour (yellow) in this analysis, as a group as one cluster.

Table IV. Subject Area.

Subject area	Documents	% of 749
Computer Science	456	60.88
Mathematics	400	53.40
Engineering	152	20.29
Physics and Astronomy	75	10.01
Biochemistry, Genetics and Molecular Biology	43	5.74
Decision Sciences	31	4.14
Neuroscience	28	3.74
Chemistry	26	3.47
Medicine	24	3.20
Materials Science	23	3.07
Agricultural and Biological Sciences	22	2.94
Multidisciplinary	22	2.94
Social Sciences	17	2.27
Environmental Science	13	1.74
Business, Management and Accounting	12	1.60
Earth and Planetary Sciences	10	1.34
Health Professions	8	1.07
Pharmacology, Toxicology and Pharmaceutics	7	0.93
Chemical Engineering	7	0.93
Arts and Humanities	4	0.53
Immunology and Microbiology	4	0.53
Energy	3	0.40

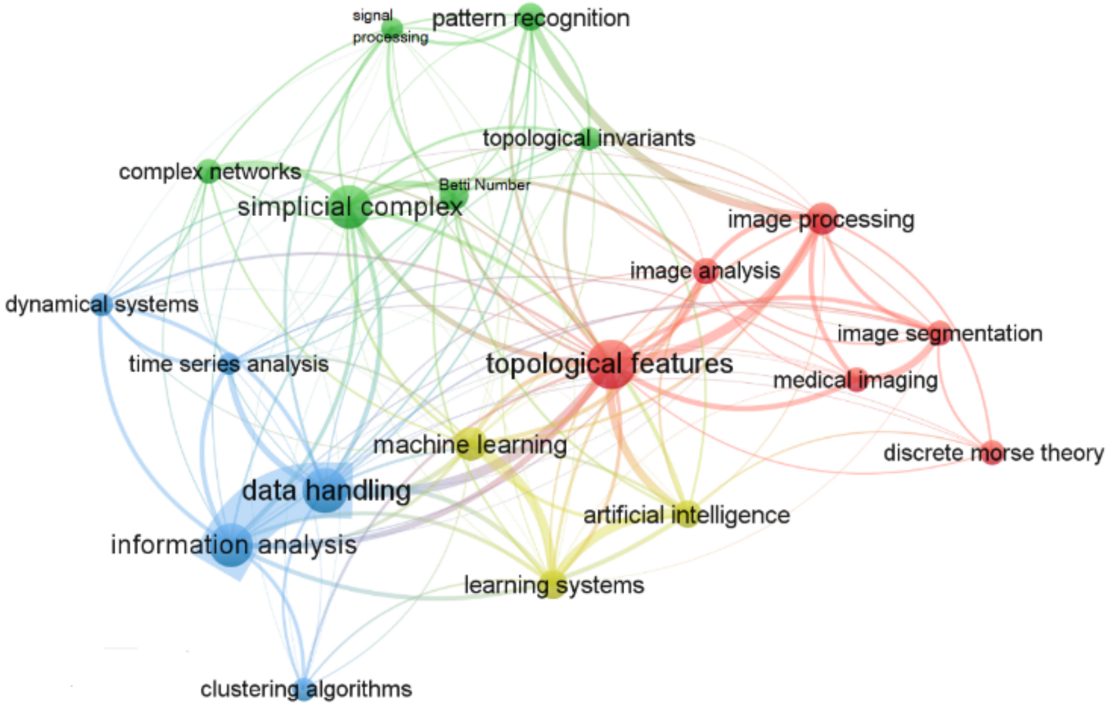

Figure 3. The network visualization map of the co-occurrence of keywords.

In the Persistent Homology research, keywords such as topological features were found as the most used keywords based on the frequency (after performing data filtration on the author's keywords). Table V exhibits the top 20 keywords (based on frequency) used in studies regarding Persistent Homology. There are 4 resulting clusters, namely cluster 1 (image analysis, cluster 2 (persistent homology interpretation), cluster 3 (time series analysis) and cluster 4 (machine learning).

Table V. Top 20 keywords in Persistent Homology

Rank	Keyword	Frequency	Cluster	DC	BC
1	topological features	81	1	19	6.40
2	data handling	70	3	18	6.22
3	simplicial complex	66	2	18	4.23
4	information analysis	63	3	18	6.12
5	machine learning	40	4	16	5.11
6	image processing	36	1	12	3.52
7	betti numbers	30	2	10	2.93
8	learning systems	30	4	9	4.45
9	pattern recognition	29	2	7	2.82
10	artificial intelligence	28	4	9	2.08
11	image analysis	26	1	12	2.34
12	discrete morse theory	24	1	6	0.31
13	medical imaging	24	1	7	0.61
14	complex networks	23	2	7	1.71
15	image segmentation	23	1	6	1.60
16	clustering algorithms	20	3	5	0.18
17	dynamical systems	20	3	6	1.09
18	topological invariants	20	2	8	1.92
19	signal processing	18	2	8	0.58
20	time series analysis	18	3	5	1.68

Based on Degree Centrality (DC), ‘topological features’ is the only node that is connected with all other nodes (19) and has a high value of Betweenness Centrality (BC). Thus, ‘topological features’ is a main and important keyword in this network. Refer to DC and BC values for other keywords, if it has a large DC value then it also has a large BC value.

Country Distribution

Overall, authors from 76 countries have contributed to the publication of documents that have been retrieved. The top 20 countries leading to Persistent Homology publications are stated in Table VI. For 392 records of publication, the US was ranked first. France (68), Italy (64), and Austria (53) followed suit. Besides total citations, the US was ranked first with a massive number of 5955 citations, followed by the UK (566) with Germany and Australia having the same number of citations (513).

Based on Degree Centrality (DC), the USA has the most connected nodes (17) followed by Austria (16) and the UK (15). Russia and Slovenia both have the lowest value of DC with only three connected nodes. Although Japan is in the fifth position of the total publication (48), the value of DC is very low (6) compared to Canada which has the same value of DC but the number of publications is two times lower than Japan.

The country with the highest betweenness is the USA. It is no coincidence that the USA has the highest number of publications and DC. Therefore, the USA is a main and important node in this network. If the USA has been removed from the network, then it will most disrupt the communications between other countries since the node is on the most information edges. Although Spain has a high DC (12), but the BC value is the lowest (0.6603). This is because the node (Spain) does not lie in is the information flow along the most geodesic (shortest path) between two vertices.

Fig. 4 further shows the authors’ network visualization map based on the countries they are affiliated with. Only countries with more than ten articles and more than five citations were considered in this analysis. Based on the fractional counting method, the finding shows that the United States plays a very prominent role in collaborating with other countries. The United States has worked closely with Israel and Denmark, while China collaborates with South Korea and Australia.

Table VI. Top 20 countries leading in Persistent Homology research

Rank	Country	Documents	Citations	DC	BC
1	USA	392	5955	17	25.62
2	France	68	823	12	10.17
3	Italy	64	733	12	6.21
4	Austria	53	420	16	15.12
5	Japan	48	285	6	6.41
6	UK	46	566	15	12.61
7	Germany	40	513	12	13.39
8	China	37	165	6	3.19
9	Spain	37	192	12	0.66
10	Canada	25	263	8	6.38
11	Australia	22	513	8	1.87
12	Poland	21	323	6	0.76
13	South Korea	20	442	7	1.28
14	Russian Federation	16	30	3	0.44
15	Israel	13	137	7	1.63
16	Singapore	11	125	4	0.5
17	Norway	10	73	4	0.36
18	Slovenia	10	143	3	0.53
19	Sweden	10	161	5	1.58
20	Switzerland	10	82	5	3.13

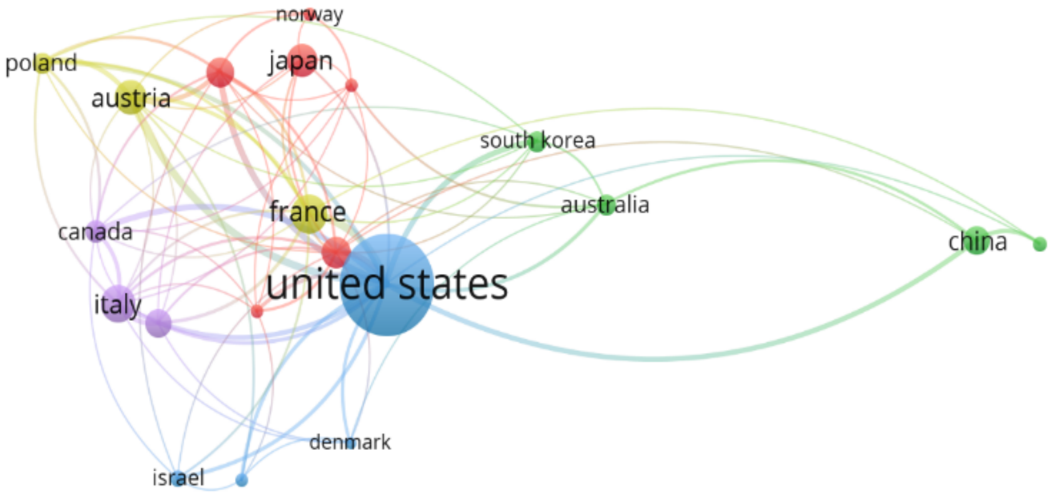

Figure 4. The country co-authorship in the field of Persistent Homology.

Authorship Analysis

This study also presents the most active authors that published the documents on Persistent Homology. Table VII listed the most active authors with a minimum of ten publications. Based on the table, Dey, T.K., Kerber, M., Edelsbrunner, H., Landi, C., Chazal, F., Wang, Y. and Carlsson, G. are among the most active authors in this field of research that publish more than 10 publications on Persistent Homology.

Based on Centrality, Wang, Y. is a main and important node (author) because he has the most connected nodes (7) and highest Betweenness Centrality (19.24). Although Dey, T.K. has the highest number of publications, he has the low value of DC and BC. Eleven authors with a zero value of BC don't have any influence because all the shortest paths between pairs of authors do not go via them.

Table VII. Most active authors in Persistent Homology research

Author	Documents	Citation	DC	BC
Dey, T.K.	21	255	3	5.29
Kerber, M.	20	258	6	12.57
Edelsbrunner, H.	19	210	3	0
Landi, C.	19	325	3	4.21
Chazal, F.	17	347	4	5.04
Wang, Y.	17	354	7	19.24
Carlsson, G.	16	1614	3	12.86
Chung, M.K.	14	347	2	4.87
Gonzalez-Diaz, R.	14	76	0	0
Hiraoka, Y.	14	188	1	0
Frosini, P.	13	256	1	0
Lee, H.	12	252	1	0
Morozov, D.	12	311	5	11.94
Sheehy, D.R.	11	116	1	0
Adams, H.	10	193	3	12.94
Bauer, U.	10	175	1	0
Boissonnat, JD.	10	73	1	0
Obayashi, I.	10	74	1	0
Petri, G.	10	258	1	0
Vaccarino, F.	10	256	1	0

This study analyses the collaboration of the authors by conducting the co-authorship analysis using VOSviewer. The research is based on influential authors with more than ten documents and has more than twenty citations—the network calculated using the fractional counting method. The nodes represent authors, edges thickness indicate the relationship's strength (numbers of document published together) amongst the authors and colour represent the cluster. Connected authors, as noted in the same colour, are commonly grouped. For example, the diagram suggests that Dey T.K., Wang Y., and Boissonnat JD were collaborating closely and usually conduct research together (see Fig 5). Fig. 5 also shows that Carlsson G. (which is coloured in red) and Kerber, M. (which is coloured in blue) have a group of authors that they collaborate with. Besides, three components of authors are not connected with the leading network.

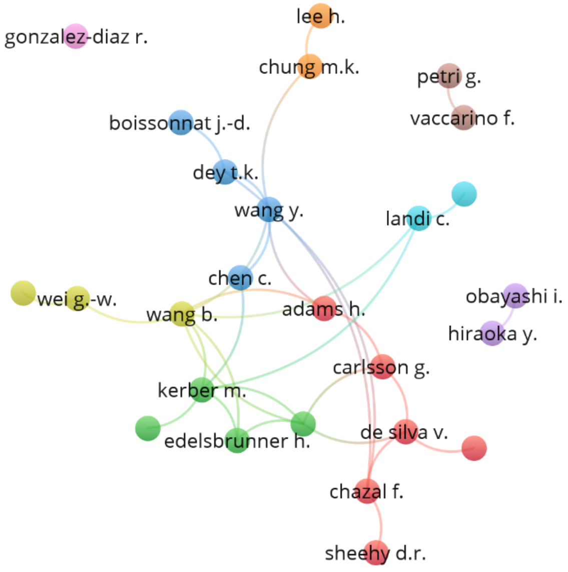

Figure 5. Co-authorship network map in the field of Persistent Homology

Further analysis for the author weighted collaboration network has been done using persistent homology. The nodes represent authors and the weight indicate the number of document published between two authors. The clique complex of the network has been obtained by filling in all cliques over multiple scales. To easily understand the concept of persistent homology in-network, see Figure 6 for an illustration of a simplicial complex, filtration, and persistent barcodes for five nodes (authors).

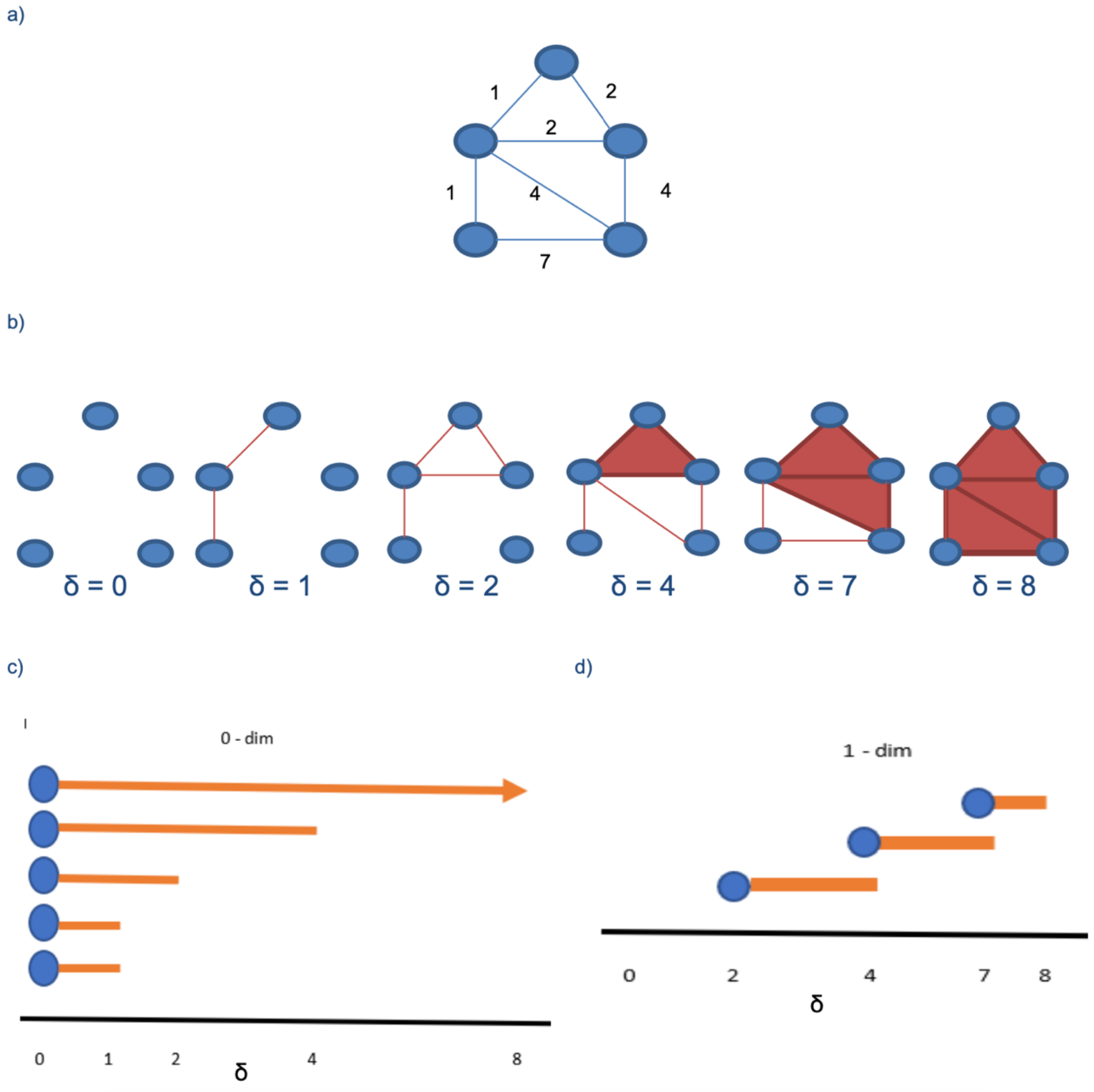

Figure 6. a) A weighted clique complex of a network. b) A filtration of the network for δ = 0,1,2,4,7,8 (from left to right). c) 0-dim Persistent barcode of the Clique complex. d) 1-dim Persistent barcode of the Clique complex.

We first explain the 0-dim of Persistent Barcodes. From Figure 6b), the filtration starts (δ = 0) with five connected components. Thus, five lines are born at the start of 0-dim Persistent barcodes (see Figure 6c)). Two components die in step two (δ = 1) since two edges are added. When δ = 2, one more edge is added, thus one component die. When δ = 4, one more component dies and is left with one connected component that persists the whole filtration.

For the 1-dim Persistent Barcodes, the first loop (hole) appears when δ = 2, thus a bar was born (see Figure 6d)). When δ = 4, the first loop is filled and die, while another loop is born. Similarly, the second loop is filled and die when δ = 7, and the last loop is born. The loop dies at δ = 8. From the 1-dim persistent barcode, we can see that there are three loops existed between five nodes.

Next, we analyze the weighted clique complex to our co-authorship network. Figure 7a shows the 0-dim and 1-dim persistent barcodes with persistent diagram of the Clique complex for the collaboration network. Due to the construction of co-authorship weighted networks, a strong link between nodes form cliques (clusters) while weak links act as bridges between those cliques, which is similar to other social networks. Based on Figure 7a, four resulting cliques survive to the end step of the filtration. One of the cliques is the leading network consisting of thirteen authors. The remaining three cliques are relatively small (with less than three authors), thus no loop is born in these cliques.

Based on Figure 7b, 1-dim persistent barcodes consist of four loops that were born in the main cliques (leading network). The colour of each bar in the 1-dim persistent barcodes refers to the clustering group in the network (see Figure 5). For example, the yellow bar is a group of three authors (Wei G.W., Wang B. and Xia K.), the red bar is a group of six authors (Adams H., Carlsson G., De Silva V., Chazal F., Sheehy D.R. and Munch E.), while each the blue bar (Boissinnat J.D., Dey T.K., Wang Y. and Chen C.) and green bar (Kerber M., Edelsbrunner H., Morozov D., Bauer U.) are the groups of 4 author. The blue, green, and red loops do not die until the last filtration, meaning that the clustering group persists and live the whole filtration. In addition, the yellow loop only has a short life since the weight (number of papers collaborate between the yellow group) is small.

Based on Figure 7c, H_0 in the persistence diagram consists of seven birth-death points even though the persistent barcode has 20 birth-death features of the connected network. This is because there are overlap points (same birth-death features) that exist. For example, there are seven components birth at $\delta = 0$ and death at $\delta = 1$. In addition, H_1 in Persistence diagram does reflect in the persistence barcode.

Citation Analysis.

We used Harzing's Publish or Perish software to screen citation metrics for data collected from the Scopus database. The data were imported into the software to construct the citation metric and total citation for each article based on Google Scholar citations. As of 6th November 2020, Table VIII provides citation indicators for records obtained. The overview contains citations per year, citations per paper, and citations per author.

Table IX summarizes the top 20 cited papers in the field of Persistent Homology. [50] received the highest citation (based on the Scopus database) with a total of 768 citations (51.2 citations per year in an article entitled "Computing persistent homology" by Carlson. The same paper has received the largest number of citations, with a total of 1283 citations and 85.83 citations per year, based on the total citations counted by Google Scholar as of 6th November 2020. [51] have the highest citation per year for both Scopus dan GS databases.

Most Active Source Title

This paper also presents the most active source title with five or more Persistent Homology publications (see Table X). Simultaneously, the information about the publisher is reported in Table X, and Springer is among the top publishers contributing to the publications on Persistent Homology, followed by Elsevier and IEEE.

Most Active Institution

The institutions' participation in the research related to persistent Homology also has been counted in this paper, i.e., based on a minimum of 15 publications. Table XI shows that Ohio State University has the highest number of publications on persistent Homology. Institute of Science and Technology Austria become the second-highest, followed by Michigan State University.

As for total citation, Stanford University has the highest number of publications on Persistent Homology. Duke University become the second-highest, followed by Michigan State University.

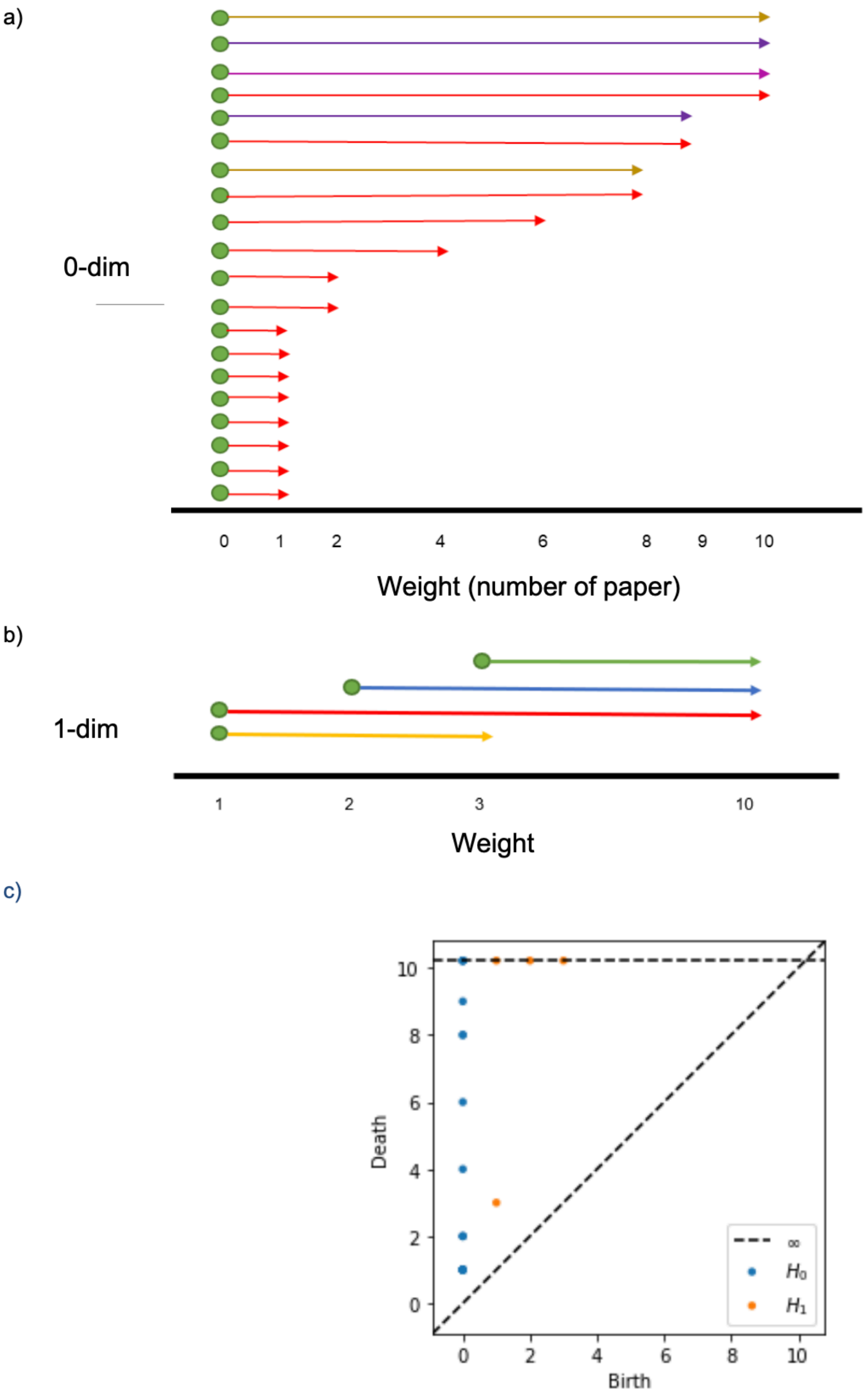

Figure 7. a) 0-dim Persistent barcode for Co-authorship network. d) 1-dim Persistent barcode for Co-authorship network. c) Persistent diagram of the Clique complex for the collaboration network. The blue dot denoted the birth-death feature of the connected network while the orange dot denoted the birth-death feature of the loop.

Table VIII. Citation Metric

Metrics	Data	Metrics	Data
Reference date:	6/11/2020	Citations/author:	3.49
Publication years:	2004-2021	Authors/paper:	3.29
Citation years:	16 (2004-2020)	Hirsch h-index:	43
Papers:	756	Egghe g-index:	72
Citations:	8669	PoP hI,norm:	22
Citations/year:	541.81	PoP hI,annual:	3.17
Citations/paper:	11.47		

Table IX. Top 20 cited papers in Persistent Homology

No.	Documents title and authors	Cited by	Cites per year	GS cites	GS cites per year
1	Computing persistent homology [50]	768	51.20	1283	85.53
2	Statistical topological data analysis using persistence landscapes [51]	269	53.80	413	82.6
3	On the local behavior of spaces of natural images [52]	267	22.25	474	39.5
4	Coverage in sensor networks via persistent Homology [53]	257	19.77	415	31.92
5	Homological scaffolds of brain functional networks [54]	205	34.17	299	49.83
6	The theory of multidimensional persistence [55]	180	16.36	369	33.55
7	Describing shapes by geometrical-topological properties of real functions [56]	156	13.00	226	18.83
8	Morse Theory for Filtrations and Efficient Computation of Persistent Homology [57]	153	21.86	243	34.71
9	Theory and algorithms for constructing discrete morse complexes from grayscale digital images [58]	146	16.22	213	23.67
10	Zigzag Persistence [59]	145	14.50	263	26.3
11	Persistence images: A stable vector representation of persistent Homology [60]	140	46.67	223	74.33
12	A roadmap for the computation of persistent Homology [4]	135	45.00	232	77.33
13	Persistent brain network homology from the perspective of dendrogram [61]	124	15.50	162	20.25
14	Topological analysis of population activity in visual corte [62]	121	10.08	179	14.92
15	Topology of viral evolution [63]	120	17.14	212	30.29
16	Confidence sets for persistence diagrams [64]	120	20.00	178	29.67
17	Hierarchical shape segmentation and registration via topological features of laplace-beltrami eigenfunctions [65]	118	11.80	171	17.1
18	A Topological Paradigm for Hippocampal Spatial Map Formation Using Persistent Homology [66]	118	14.75	159	19.88
19	Towards persistence-based reconstruction in euclidean spaces [67]	106	8.83	169	14.08
20	Zigzag persistent Homology and real-valued functions [68]	95	8.64	207	18.82

Table X. Most active source title.

	Source Title	Publisher	Document	Citation
1	Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics	Springer	67	567
2	Leibniz International Proceedings In Informatics Lipics	Dagstuhl Publishing	30	66
3	Proceedings Of The Annual Symposium On Computational Geometry	Association for Computing Machinery	20	495
4	Discrete And Computational Geometry	Springer	17	1059
5	Foundations Of Computational Mathematics	Springer-Verlag	16	378
6	Pattern Recognition Letters	Elsevier	13	81
7	Proceedings International Symposium On Biomedical Imaging	IEEE	10	66
8	Proceedings Of The Annual ACM SIAM Symposium On Discrete Algorithms	SIAM	10	73
9	ICASSP IEEE International Conference On Acoustics Speech And Signal Processing Proceedings	IEEE	9	50
10	Physical Review ES	APS	8	65
11	Proceedings 18th IEEE International Conference On Machine Learning And Applications Icmla 2019	IEEE	8	0
12	Scientific Reports	Nature Research	8	43
13	Computational Geometry Theory And Applications	Elsevier	7	98
14	Homology Homotopy And Applications	International Press	7	94
15	Mathematics And Visualization	Springer	7	73
16	Physica D Nonlinear Phenomena	Elsevier	7	82
17	IEEE Transactions On Visualization And Computer Graphics	IEEE	6	112
18	SIAM Journal On Applied Algebra And Geometry	SIAM	6	14
19	Abel Symposia	Springer	5	0
20	Advances In Intelligent Systems And Computing	Springer	5	2
21	Computer Graphics Forum	Wiley	5	105
22	Computer Vision And Image Understanding	Elsevier	5	59
23	IEEE Computer Society Conference On Computer Vision And Pattern Recognition Workshops	IEEE	5	83
24	Journal Of Machine Learning Research	JMLR, Inc. and Microtome Publishing	5	287

Table XI. Most active institution

Affiliation	Total Publication	Total Citation
The Ohio State University	44	441
Institute of Science and Technology Austria	33	461
Michigan State University	29	561
Duke University	28	935
Stanford University	28	2531
Alma Mater Studiorum Università di Bologna	25	432
University of Seville	24	139
INRIA Saclay	24	451
Università degli Studi di Modena e Reggio Emilia	20	403
University of Pennsylvania	20	396
University of Wisconsin-Madison	20	475
INRIA Sophia Antipolis	19	158
Uniwersytet Jagielloński w Krakowie	18	284
Tohoku University	15	263
Technische Universitat Graz	15	135
University of Minnesota Twin Cities	15	323

Discussion and Conclusion

This study aims to examine the trend of research on Persistent Homology using bibliometric analysis. After data filtration, 756 documents were extracted from the Scopus database. The research on Persistent Homology started by [9], and the number of published papers increased year by year. Based on the 17 years of publications (2004-2020), there are 8669 citations with 541.81 citations per year. While there is drastic progress on the topic for the past five years, we believed that research related to Persistent Homology has a brilliant opportunity for researchers to study and produce more quality research papers.

Since Persistent Homology is more focused on studies related to computer science and mathematics, it can be found that both subject areas were representing 60.88% and 53.40% of the total publications, respectively. This study also reveals that the areas mostly covered in the Persistent Homology research are related to "topological features" and "data handling" based on the author's keywords. Other potential topics worth exploring include "machine learning," "image segmentation," and "time series analysis."

Despite valuable insights offered by this article, readers should consider several limitations. First, the results only emerged from the specific keyword based on the document's title and keyword. The search query results on the other fields, such as abstract, have not been included in this analysis. Future research probably can be expanded into the abstract field. However, data cleaning or data filtering should be done first based on the possibility that Persistent Homology is only referred to as a recommendation and not the main research work.

It is also worth noting that no search query is 100% perfect; thus, false positive and negative results should be anticipated [69]. Thirdly, this research is based only on the Scopus database. Other available databases such as Web of Science, Google Scholar and even unindexed journals probably can be included in the future bibliometric study. Despite these limitations, our research was the first to analyze the bibliometric analysis of Persistent Homology literature.

References

[1] Y. Du, M. Zhang, G. Stonis, and S. Juan, "Topological Data Analysis on Magnetic Resonance Image Biomarkers," 2019, pp. 1185–1187, doi: 10.1109/BIBM47256.2019.8983312.

[2] M. Olejniczak, A. Severo Pereira Gomes, and J. Tierny, "A Topological Data Analysis perspective on noncovalent interactions in relativistic calculations," *Int. J. Quantum Chem.*, vol. 120, no. 8, p. e26133, Apr. 2020, doi: https://doi.org/10.1002/qua.26133.

[3] S. Agami and R. J. Adler, "Modeling of persistent homology," *Commun. Stat. - Theory Methods*, vol. 49, no. 20, pp. 4871–4888, 2020, doi: 10.1080/03610926.2019.1615091.

[4] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington, "A roadmap for the computation of persistent homology," *EPJ Data Sci.*, vol. 6, no. 1, pp. 1–45, 2017, doi: 10.1140/epjds/s13688-017-0109-5.

[5] K. Turner, S. Mukherjee, and D. M. Boyer, "Persistent homology transform for modeling shapes and surfaces," *Inf. Inference*, vol. 3, no. 4, pp. 310–344, 2014, doi: 10.1093/imaiai/iau011.

[6] D. Bhaskar *et al.*, "Analyzing collective motion with machine learning and topology," *Chaos*, vol. 29, no. 12, pp. 123125-1-12315–12, 2019, doi: 10.1063/1.5125493.

[7] M. E. Aktas, E. Akbas, and A. El Fatmaoui, "Persistence homology of networks: methods and applications," *Appl. Netw. Sci.*, vol. 4, no. 1, pp. 1–28, 2019, doi: 10.1007/s41109-019-0179-3.

[8] M. Offroy and L. Duponchel, "Topological data analysis: A promising big data exploration tool in biology, analytical chemistry and physical chemistry," *Anal. Chim. Acta*, vol. 910, pp. 1–11, 2016, doi: 10.1016/j.aca.2015.12.037.

[9] A. Zomorodian and G. Carlsson, "Computing persistent homology," in *Proceedings of the Twentieth Annual Symposium on Computational Geometry (SCG'04)*, 2004, pp. 347–356, doi: 10.1145/997817.997870.

[10] C. S. Pun, K. Xia, and S. X. Lee, "Persistent-Homology-Based Machine Learning and Its Applications -- A Survey," *SSRN Electron. J.*, 2018, doi: 10.2139/ssrn.3275996.

[11] F. Altındiş, B. Yılmaz, S. Borisenok, and K. İçöz, "Parameter investigation of topological data analysis for EEG signals," *Biomed. Signal Process. Control*, vol. 63, 2021, doi: 10.1016/j.bspc.2020.102196.

[12] Y. Wang, H. Ombao, and M. K. Chung, "Statistical Persistent Homology of Brain Signals," 2019, vol. 2019-May, pp. 1125–1129, doi: 10.1109/ICASSP.2019.8682978.

[13] E. Merelli, M. Piangerelli, M. Rucco, and D. Toller, "A topological approach for multivariate time series characterization: The epileptic brain," 2015, doi: 10.4108/eai.3-12-2015.2262525.

[14] S. Majumder, F. Apicella, F. Muratori, and K. Das, "Detecting autism spectrum disorder using topological data analysis," 2020, vol. 2020-May, pp. 1210–1214, doi: 10.1109/ICASSP40776.2020.9054747.

[15] F. Nasrin, C. Oballe, D. L. Boothe, and V. Maroulas, "Bayesian topological learning for brain state classification," 2019, pp. 1247–1252, doi: 10.1109/ICMLA.2019.00205.

[16] Y. Wang, H. Ombao, and M. K. Chung, "Topological seizure origin detection in electroencephalographic signals," 2015, vol. 2015-July, pp. 351–354, doi: 10.1109/ISBI.2015.7163885.

[17] Y. Wang, H. Ombao, and M. K. Chung, "Topological data analysis of single-trial electroencephalographic signals," *Ann. Appl. Stat.*, vol. 12, no. 3, pp. 1506–1534, 2018, doi: 10.1214/17-AOAS1119.

[18] Y. Wang, R. Behroozmand, L. P. Johnson, and J. Fridriksson, "Topology Highlights Neural Deficits of Post-Stroke Aphasia Patients," 2020, vol. 2020-April, pp. 754–757, doi: 10.1109/ISBI45749.2020.9098734.

[19] M. Nguyen, M. Aktas, and E. Akbas, "Bot Detection on Social Networks Using Persistent Homology," *Math. Comput. Appl.*, vol. 25, no. 3, pp. 1–16, 2020, doi: 10.3390/mca25030058.

[20] S. Chowdhury and F. Mémoli, "Persistent path homology of directed networks," 2018, pp. 1152–1169, doi: 10.1137/1.9781611975031.75.

[21] T. Gao and F. Li, "Studying the utility preservation in social network anonymization via persistent homology," *Comput. Secur.*, vol. 77, pp. 49–64, 2018, doi: 10.1016/j.cose.2018.04.003.

[22] X. Wang, M. Bennamoun, F. Sohel, and H. Lei, "Diffusion geometry derived keypoints and local descriptors for 3D deformable shape analysis," *J. Circuits, Syst. Comput.*, 2020, doi: 10.1142/S021812662150016X.

[23] H. Edelsbrunner, "Shape, homology, persistence, and stability," vol. 9411. Springer Verlag, Institute of Science and Technology, Austria, 2015, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84951996567&partnerID=40&md5=31099c8c3f5378ccf2aa9465eb46773c.

[24] P. Frosini and G. Jabłoński, "Combining Persistent Homology and Invariance Groups for Shape Comparison," *Discret. Comput. Geom.*, vol. 55, no. 2, pp. 373–409, 2016, doi: 10.1007/s00454-016-9761-y.

[25] X. Wang, F. Sohel, M. Bennamoun, Y. Guo, and H. Lei, "Scale space clustering evolution for salient region detection on 3D deformable shapes," *Pattern Recognit.*, vol. 71, pp. 414–427, 2017, doi: 10.1016/j.patcog.2017.05.018.

[26] C. M. Topaz, L. Ziegelmeier, and T. Halverson, "Topological data analysis of biological aggregation models," *PLoS One*, vol. 10, no. 5, pp. 1–25, 2015, doi: 10.1371/journal.pone.0126383.

[27] M. J. Jimenez, M. Rucco, P. Vicente-Munuera, P. Gómez-Gálvez, and L. M. Escudero, "Topological data analysis for self-organization of biological tissues," vol. 10256 LNCS. Springer Verlag, Departamento Matematica Aplicada I, Universidad de Sevilla, Campus Reina Mercedes, Sevilla, 41012, Spain, pp. 229–242, 2017, doi: 10.1007/978-3-319-59108-7_18.

[28] C. S. Pun, B. Y. Sin Yong, and K. Xia, "Weighted-persistent-homology-based machine learning for RNA flexibility analysis," *PLoS One*, vol. 15, no. 8 August 2020, 2020, doi: 10.1371/journal.pone.0237747.

[29] S. N. Chukanov, "Comparison of objects' images based on computational topology methods ," *SPIIRAS Proc.*, vol. 18, no. 5, pp. 1043–1065, 2019, doi: 10.15622/sp.2019.18.5.1043-1065.

[30] O. Ege and I. Karaca, "Persistent homology of graph-like digital images," *Ann. di Mat. Pura ed Appl.*, vol. 199, no. 6, pp. 2167–2179, 2020, doi: 10.1007/s10231-020-00962-x.

[31] O. Dunaeva, H. Edelsbrunner, A. Lukyanov, M. Machin, and D. Malkova, "The classification of endoscopy images with persistent homology," 2015, pp. 565–570, doi: 10.1109/SYNASC.2014.81.

[32] K. Watarai, H. Omura, and T. Minamoto, "Esophageal Abnormality Detection from Endoscopic Images Using DT-CDWT and Persistent Homology," vol. 1134. Springer, Graduate School of Science and Engineering of Saga University, Saga, Japan, pp. 227–232, 2020, doi: 10.1007/978-3-030-43020-7_30.

[33] J. R. Tempelman and F. A. Khasawneh, "A look into chaos detection through topological data analysis," *Phys. D Nonlinear Phenom.*, vol. 406, 2020, doi: 10.1016/j.physd.2020.132446.

[34] N. F. S. Zulkepli, M. S. M. Noorani, F. A. Razak, M. Ismail, and M. A. Alias, "Topological characterization of haze episodes using persistent homology," *Aerosol Air Qual. Res.*, vol. 19, no. 7, pp. 1614–1621, 2019, doi: 10.4209/aaqr.2018.08.0315.

[35] N. F. S. Zulkepli, M. S. Md Noorani, F. A. Razak, M. Ismail, and M. A. Alias, "Cluster analysis of haze episodes based on topological features," *Sustain.*, vol. 12, no. 10, 2020, doi: 10.3390/SU12103985.

[36] N. F. S. Zulkepli, M. S. M. Noorani, F. A. Razak, M. Ismail, and M. A. Alias, "Pendekatan Baharu untuk Mengelompok Stesen Pengawasan Kualiti Udara menggunakan Homologi Gigih," *Sains Malaysiana*, vol. 49, no. 4, pp. 963–970, 2020, doi: 10.17576/jsm-2020-4904-25.

[37] M. Gidea, D. Goldsmith, Y. Katz, P. Roldan, and Y. Shmalo, "Topological recognition of critical transitions in time series of cryptocurrencies," *Phys. A Stat. Mech. its Appl.*, vol. 548, p. 123843, 2020, doi: 10.1016/j.physa.2019.123843.

[38] M. Gidea and Y. Katz, "Topological data analysis of financial time series: Landscapes of crashes," *Phys. A Stat. Mech. its Appl.*, vol. 491, pp. 820–834, 2018, doi: 10.1016/j.physa.2017.09.028.

[39] M. S. Ismail, M. S. Md Noorani, M. Ismail, F. Abdul Razak, and M. A. Alias, "Predicting next day direction of stock price movement using machine learning methods with persistent homology: Evidence from Kuala Lumpur Stock Exchange," *Applied Soft Computing Journal*, vol. 93. 2020, doi: 10.1016/j.asoc.2020.106422.

[40] A. Ahmi, H. Elbardan, and R. H. Raja Mohd Ali, "Bibliometric analysis of published literature on industry 4.0," *ICEIC 2019 - Int. Conf. Electron. Information, Commun.*, vol. 0, pp. 1–6, 2019, doi: 10.23919/ELINFOCOM.2019.8706445.

[41] A. Pritchard, "Statistical Bibliography or Bibliometrics," *J. Doc.*, vol. 25, pp. 348–349, 1969.

[42] S. B. Roy and M. Basak, "Journal of documentation: A bibliometric study," *Libr. Philos. Pract.*, vol. 2013, no. August, 2013.

[43] A. W. Harzing, "Publish or Perish," 2007. https://harzing.com/resources/publish-or-perish.

[44] N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010, doi: 10.1007/s11192-009-0146-3.

[45] S. Zaoli, P. Mazzarisi, and F. Lillo, "Betweenness centrality for temporal multiplexes," *Sci. Rep.*, vol. 11, no. 1, pp. 1–9, 2021, doi: 10.1038/s41598-021-84418-z.

[46] J. Zhang and Y. Luo, "Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network," vol. 132, no. Msam, pp. 300–303, 2017, doi: 10.2991/msam-17.2017.68.

[47] N. M. Rasidi, S. A. Bakar, and F. A. Razak, "Modelling the correlation of PM10 concentration and location of air quality monitoring stations in Malaysia using network method," *AIP Conf. Proc.*, vol. 1750, no. 2016, 2016, doi: 10.1063/1.4954560.

[48] C. J. Carstens and K. J. Horadam, "Persistent homology of collaboration networks," *Math. Probl. Eng.*, vol. 2013, 2013, doi: 10.1155/2013/815035.

[49] N. Saul and C. Tralie, "Scikit-TDA: Topological Data Analysis for Python," 2019.

[50] A. Zomorodian and G. Carlsson, "Computing persistent homology," *Discret. Comput. Geom.*, vol. 33, no. 2, pp. 249–274, 2005, doi: 10.1007/s00454-004-1146-y.

[51] P. Bubenik, "Statistical topological data analysis using persistence landscapes," *J. Mach. Learn. Res.*, vol. 16, pp. 77–102, 2015, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84923901980&partnerID=40&md5=526178c05d43c6b74894dba6c69d5579.

[52] G. Carlsson, T. Ishkhanov, V. De Silva, and A. Zomorodian, "On the local behavior of spaces of natural images," *Int. J. Comput. Vis.*, vol. 76, no. 1, pp. 1–12, 2008, doi: 10.1007/s11263-007-0056-x.

[53] V. de Silva and R. Ghrist, "Coverage in sensor networks via persistent homology," *Algebr. Geom. Topol.*, vol. 7, no. 1, pp. 339–358, 2007, doi: 10.2140/agt.2007.7.339.

[54] G. Petri *et al.*, "Homological scaffolds of brain functional networks," *J. R. Soc. Interface*, vol. 11, no. 101, 2014, doi: 10.1098/rsif.2014.0873.

[55] G. Carlsson and A. Zomorodian, "The theory of multidimensional persistence," *Discret. Comput. Geom.*, vol. 42, no. 1, pp. 71–93, 2009, doi: 10.1007/s00454-009-9176-0.

[56] S. Biasotti *et al.*, "Describing shapes by geometrical-topological properties of real functions," *ACM Comput. Surv.*, vol. 40, no. 4, 2008, doi: 10.1145/1391729.1391731.

[57] K. Mischaikow and V. Nanda, "Morse Theory for Filtrations and Efficient Computation of Persistent Homology," *Discret. Comput. Geom.*, vol. 50, no. 2, pp. 330–353, 2013, doi: 10.1007/s00454-013-9529-6.

[58] V. Robins, P. J. Wood, and A. P. Sheppard, "Theory and algorithms for constructing discrete morse complexes from grayscale digital images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1646–1658, 2011, doi: 10.1109/TPAMI.2011.95.

[59] G. Carlsson and V. de Silva, "Zigzag Persistence," *Found. Comput. Math.*, vol. 10, no. 4, pp. 367–405, 2010, doi: 10.1007/s10208-010-9066-0.

[60] H. Adams *et al.*, "Persistence images: A stable vector representation of persistent homology," *J. Mach. Learn. Res.*, vol. 18, pp. 1–35, 2017, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85014704326&partnerID=40&md5=6ec341f4c78c88c362d6bc308942df1e.

[61] H. Lee, H. Kang, M. K. Chung, B.-N. Kim, and D. S. Lee, "Persistent brain network homology from the perspective of dendrogram," *IEEE Trans. Med. Imaging*, vol. 31, no. 12, pp. 2267–2277, 2012, doi: 10.1109/TMI.2012.2219590.

[62] G. Singh, F. Memoli, T. Ishkhanov, G. Sapiro, G. Carlsson, and D. L. Ringach, "Topological analysis of population activity in visual cortex," *J. Vis.*, vol. 8, no. 8, 2008, doi: 10.1167/8.8.11.

[63] J. M. Chan, G. Carlsson, and R. Rabadan, "Topology of viral evolution," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 46, pp. 18566–18571, 2013, doi: 10.1073/pnas.1313480110.

[64] B. T. Fasy, F. Lecci, A. Rinaldo, L. Wasserman, S. Balakrishnan, and A. Singh, "Confidence sets for persistence diagrams," *Ann. Stat.*, vol. 42, no. 6, pp. 2301–2339, 2014, doi: 10.1214/14-AOS1252.

[65] M. Reuter, "Hierarchical shape segmentation and registration via topological features of laplace-beltrami eigenfunctions," *Int. J. Comput. Vis.*, vol. 89, no. 2–3, pp. 287–308, 2010, doi: 10.1007/s11263-009-0278-1.

[66] Y. Dabaghian, F. Mémoli, L. Frank, and G. Carlsson, "A Topological Paradigm for Hippocampal Spatial Map Formation Using Persistent Homology," *PLoS Comput. Biol.*, vol. 8, no. 8, 2012, doi: 10.1371/journal.pcbi.1002581.

[67] F. Chazal and S. Y. Oudot, "Towards persistence-based reconstruction in euclidean spaces," 2008, pp. 232–241, doi: 10.1145/1377676.1377719.

[68] G. Carlsson, V. De Silva, and D. Morozov, "Zigzag persistent homology and real-valued functions," 2009, pp. 247–256, doi: 10.1145/1542362.1542408.

[69] A. Ahmi and M. H. Mohd Nasir, "Examining the trend of the research on extensible business reporting language (xbrl): A bibliometric review," *Int. J. Innov. Creat. Chang.*, vol. 5, no. 2, 2019.