Journal of Theoretical and Applied Information Technology

15th November 2023. Vol.101. No 21
© 2023 Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MAPPING THE PHISHING ATTACKS RESEARCH LANDSCAPE: A BIBLIOMETRIC ANALYSIS AND TAXONOMY

MELTEM MUTLUTÜRK[1] , BILGIN METIN[1]

[1]Department of Management Information Systems, Bogazici University, Istanbul, Turkey

E-mail: meltem.mutluturk@boun.edu.tr, bilgin.metin@boun.edu.tr

ABSTRACT

Phishing attacks represent a worldwide issue that requires a comprehensive, global strategy to tackle. Delving into international scholarly research on phishing incidents allows us to grasp the extent of this problem on a global scale, while taking into account the unique obstacles and perspectives that arise in different regions. This bibliometric study offers a comprehensive analysis of the phishing research domain from 2004 to 2023, highlighting the growth, trends, and collaborative networks shaping this field. The presented study uncovers the most influential articles, authors, and institutions, as well as emerging research themes and collaborative patterns using network analyses, including citation, co-citation, co-authorship, co-occurrence, and bibliographic coupling. The results demonstrate a consistent growth in the number of publications, indicating the increased interest and relevance of phishing research in addressing cybersecurity challenges. The study identifies the main research clusters and emerging topics, offering insights into future research directions and practical applications. Furthermore, the analysis emphasizes the importance of fostering interdisciplinary collaboration and academia-industry-government partnerships to develop more effective countermeasures against phishing attacks. By understanding the current research landscape and promoting stronger partnerships, stakeholders can work together to devise innovative strategies and tools to protect individuals and organizations from phishing threats. Lastly, the study provides a taxonomy of the phishing literature.

Keywords: *Phishing, Bibliometric, Taxonomy, Vosviewer, Collaboration*

1. INTRODUCTION

Phishing has emerged as a pervasive cybersecurity threat in recent years, posing significant risks to individuals, businesses, and organizations worldwide. Phishing attacks involve deceptive communication, typically via email or other electronic messaging platforms, to manipulate users into disclosing sensitive information, such as login credentials or financial data, or executing malicious actions under the guise of a trustworthy entity or individual [1]. During the third quarter of 2022, the APWG reported a record-breaking 1,270,883 phishing attacks, marking the worst quarter for phishing incidents ever documented by the organization [2]. The increasing prevalence of phishing has prompted growing concern and a heightened focus on understanding the factors contributing to its success, as well as developing effective countermeasures to mitigate its impact.

Recent studies have examined various aspects of phishing, including the development of machine learning-based detection and prevention systems [3], the role of human factors in phishing susceptibility [4], [5], [6] and the effectiveness of anti-phishing education and training initiatives [7]. As phishing attacks continue to evolve in sophistication and scope, a thorough understanding of the current state of research in this domain is vital for guiding future investigations and ensuring the security of digital environments.

Bibliometric analyses have been widely employed to assess the research productivity and impact of various scientific disciplines [8]. However, research on bibliometric analysis of phishing attacks remains limited, predominantly concentrating on individual industries or general cybersecurity threats [9]. To the best of the authors' knowledge, the topic of phishing attacks remains relatively unexplored, utilizing visual representations created by VOSviewer. While numerous studies have explored the menace of phishing attacks, few have provided a structured and analytical overview of the collective research in the domain. Previous bibliometric studies on phishing have been restricted to either individual sectors or focused on broad cybersecurity themes, leaving a knowledge gap in understanding the specific landscape of phishing research. Our motivation stems from this gap, aiming to offer an in-depth bibliometric scrutiny that not only synthesizes the current state of knowledge but also lays the foundation for future inquiries. This study presents a bibliometric analysis of the phishing research landscape, drawing on articles indexed in the Web of Science database. This study contributes to the existing literature by providing a holistic overview of the phishing research domain, enabling scholars and practitioners to better understand the current state of research and potential avenues for future investigation. By analysing the citations, co-authorship patterns, and keyword co-occurrence networks, this study aims to uncover the intellectual structure and evolution of phishing research, identify influential authors, institutions, and countries, and uncover emerging research fronts.

The main objectives of this study are to:

1. Provide a holistic bibliometric analysis of international research on phishing attacks, with a particular focus on the scope and regional differences.

2. Identify prevailing themes, trends, and knowledge gaps in phishing research literature.

3. Examine publication patterns, collaboration networks, and ascertain the impact of seminal articles and authors in this research area.

4. Track the growth trajectory and evolution of research focusing on phishing attacks and related human factors.

5. Propose a taxonomy of phishing attacks derived from keyword occurrences in the analysed articles.

This study does not delve into the technical intricacies of phishing attacks or provide primary empirical research. Instead, it focuses on offering a structured review and bibliometric analysis of existing literature.

2. RELATED WORK

Phishing attacks have been a prevalent issue in cybersecurity, with various studies aiming to define and classify these attacks. These attacks often involve electronic communications (e.g., email, SMS, VOIP, instant messaging) that seem to come from reliable entities [1] in order to deceive users into clicking malicious links or downloading harmful files [10]. Phishing attacks typically progress through three phases: (1) evading technical cybersecurity defences to deliver the deceptive message to the target, (2) persuading the target to follow the suggested action, and (3) the attacker leveraging the delivered payload for personal gain. These gains can take various forms, including financial losses, espionage, theft of trade secrets, and sabotage, as those behind phishing attacks have diverse malicious intentions [11]. Types of phishing attacks have been categorized based on their targets and techniques, such as spear phishing, whaling, and smishing [12]. Researchers have also developed taxonomies for phishing attacks to better understand their delivery methods, targets, and goals [13], [14], [11], [12]. These taxonomies are essential for providing a foundation for further research and the development of effective countermeasures.

A significant body of research has focused on the detection and prevention of phishing attacks. Machine learning and data mining techniques have been employed to detect phishing websites and emails, with promising results in terms of accuracy and efficiency [15], [16], [17], [18]. Additionally, human factors have been investigated, with studies examining user awareness and education as essential elements in preventing phishing attacks [19], [20], [21], [22]. Despite these efforts, phishing attacks continue to evolve and adapt, necessitating further research into novel detection and prevention techniques.

While bibliometric analyses have been conducted on broader cybersecurity topics, focused analyses on phishing research are limited. For example, [23] conducted a bibliometric analysis on information security in business, revealing key trends, influential publications, and research gaps. However, a bibliometric analysis specifically addressing phishing research is necessary to understand the current state of the literature, identify research gaps, and highlight the most impactful studies in this field.

In summary, the literature on phishing attacks encompasses various aspects, including regional differences, definitions and types of phishing attacks, techniques and strategies employed by attackers, detection and prevention methods, bibliometric analysis of phishing research, and the impact of phishing on individuals and organizations. This study aims to contribute to the body of knowledge by conducting a bibliometric analysis of phishing research, identifying emerging trends, and highlighting potential research gaps that could inform future studies and the development of effective countermeasures against phishing threats.

This study seeks to address the existing gap in the literature by performing a bibliometric analysis of phishing attack research. The objectives of this investigation include:

1. Examining international research on phishing attacks could be beneficial to understand the issue's scope and regional differences since phishing attacks are a global concern and require a holistic approach.

2. Identifying the key themes, trends, and gaps in the literature surrounding phishing attacks.

3. Examining the publication patterns, collaboration networks, and impact of individual articles and authors within this research area.

4. Evaluating the growth and evolution of research on phishing attacks and human factors over time.5. Developing a taxonomy of phishing based on keyword occurrences in the analysed articles.

By accomplishing these objectives, the study will enrich the existing literature by offering a comprehensive overview of the current state of knowledge on phishing attacks.

3. METHODOLOGY

3.1 Bibliometric Analysis

Bibliometric analysis is a prevalent technique in scientometrics, which entails the quantitative assessment of scientific literature by evaluating publication patterns, citation networks, and collaborative connections between researchers and institutions. This method allows researchers to map the intellectual structure, identify main themes and trends, and evaluate the impact of individual publications and authors within a specific area. One of the main advantages of bibliometric studies is their ability to provide a comprehensive, impartial, and data-driven overview of a research field, helping researchers and policymakers identify research gaps, emerging areas, and potential avenues for future investigations. The bibliometric approach involves employing quantitative methods, such as bibliometric analysis (e.g., citation analysis), on bibliometric data, which includes publication and citation data units [24].

Scientometric techniques and tools are utilized by scholars to measure the research productivity of other scientists, forecast their future career trajectories, and assess the effects of funding decisions on the academic community's structure. These techniques depend on academic bibliographic data and essential scientometric tools to create knowledge domain maps [25].

In the context of phishing and human factors, a bibliometric analysis is particularly valuable, as it can help consolidate the growing literature in this area and provide a systematic understanding of the current state of knowledge. Furthermore, it can reveal the extent of research collaboration and interdisciplinary ties between cybersecurity, human behaviour, psychology, and other related fields, encouraging interdisciplinary research.

Science mapping examines the relationships among research components [26]. This analysis investigates the intellectual exchanges and structural connections between these elements. Various science mapping methods include citation analysis, co-citation analysis, bibliographic coupling, co-word analysis, and co-authorship analysis [27].

3.2 Data Collection

For the bibliometric analysis, the Web of Science (WoS) database served as the primary source for obtaining pertinent publications. WoS was chosen for several reasons: (1) it is a well-recognized and all-encompassing repository for scientific literature [28], covering a wide variety of research fields, such as cybersecurity and human behavior; (2) it provides extensive citation information [29], allowing for the exploration of citation networks and evaluation of the impact of individual publications and authors; and (3) its advanced search and filtering options make it easier to efficiently find relevant articles based on keywords, publication years, and other factors.

A search query was created using only the keyword "phishing" as this research aims to explore all research on this topic. This keyword search resulted in 3139 articles.

For this bibliometric analysis, articles were selected based on the following criteria:

1. Indexed in the Web of Science database.
2. Primary focus on phishing attacks.
3. Written in English.
4. Excludes short communications, commentaries, and editorials.

The above criteria were chosen to ensure comprehensive coverage of seminal and recent research in the domain, while maintaining a manageable and relevant dataset for analysis.

3.3 Data Analysis and Visualization

For data analysis and visualization, VOSviewer, a widely used software tool designed to build and visualize bibliometric networks, was employed. VOSviewer enables the creation of various network types, such as co-authorship, co-citation, and keyword co-occurrence networks, utilizing advanced clustering algorithms [30] to uncover the underlying structure and primary themes within the research domain. Additionally, it offers an interactive and user-friendly interface for examining and interpreting the visualized networks.

Using VOSviewer, networks were created and analysed based on publication and citation data, a collaboration between authors and institutions, and keyword co-occurrence. The outcomes were visualized as network maps, emphasizing crucial clusters, trends, and connections within the phishing domain.

4. RESULTS

4.1 Publications

Figure 1 provides a visual representation of the growth and trends in the phishing research domain from 2004 to 2023. The number of articles published on the subject of phishing demonstrates a consistent upward trajectory, indicating increased interest and attention from researchers and the academic community. However, the total citations show some fluctuations, with a general increase from 2004 to 2019, followed by a decline in 2020 and beyond. The drop in citation counts could be influenced by several factors, such as the time lag between publication and citation, changes in research focus, or the emergence of new, related research areas [31]. It could also be attributed to the fact that over a three-year period, interdisciplinary papers tend to accumulate fewer citations than the average, but this trend reverses over a 13-year period. Papers combining vastly different fields often attract fewer citations [32]. The incomplete data for 2023 also contributes to the observed dip in citation counts for that year.

It is important to note that the steady increase in the number of phishing-related articles reflects the growing importance of this research area, as phishing attacks continue to evolve and pose significant challenges to individuals, organizations, and society as a whole [15], [22].

The emergence of articles on the topic of phishing in 2004 can be attributed to the increasing prominence and recognition of phishing attacks as a significant cybersecurity threat during that time [33], [12]. The year 2004 marks the beginning of this growing interest in phishing research, as evidenced by the publication of scholarly articles addressing various aspects of phishing.

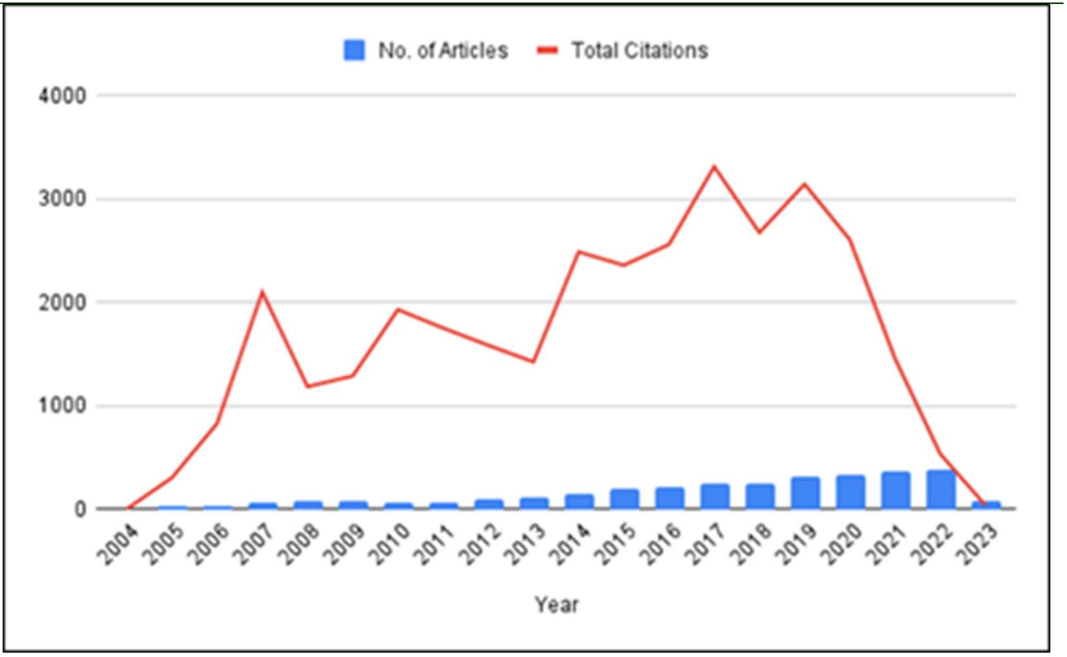

Figure. 1.Frequency of publications and corresponding citations.

In total, the 3139 articles gathered 33595 citations, with an average of 10.7 citations per article. Table 1 shows the yearly count of articles, total citations, and average citations per article, with Figure 1 displaying the peak in citations in 2017. However, 2007 has the highest average citations per article.

Table 1: Frequency of Articles, Total Citations, and Citations Per Article

Year	No. of Articles	Total Citations	CPA
2004	2	1	0.5
2005	21	303	14.4
2006	36	829	23.0
2007	68	2101	30.9
2008	79	1186	15.0
2009	83	1288	15.5
2010	65	1931	29.7
2011	67	1753	26.2
2012	89	1586	17.8
2013	108	1425	13.2
2014	152	2490	16.4
2015	198	2361	11.9
2016	222	2562	11.5
2017	247	3316	13.4
2018	254	2677	10.5
2019	314	3146	10.0
2020	331	2609	7.9
2021	357	1459	4.1
2022	375	535	1.4
2023	71	37	0.5

In order to effectively analyse the topics and types of articles within each period, the most cited articles have been divided into two distinct 10-year intervals, as employed in the study by [34]. This approach facilitates the examination of more recent research, while taking into account the citation gap that exists between older and newer articles. Table 2 presents the top ten articles from the first period (2004–2013), while Table 3 details the top ten articles from the second period (2014–2023).

The most cited article in the first period [33] aimed to ethically quantify how reliable social context would increase the success of a phishing attack, targeting Indiana University students aged 18 to 24 years old. The findings indicated a need for extensive educational campaigns about phishing and other security threats to raise awareness and reduce vulnerabilities among internet users whereas the most cited article in the second period [35] investigates the growing threat of social engineering in virtual communities, particularly due to the increasing adoption of Bring Your Own Device (BYOD) policies and online communication tools. It offers a taxonomy of well-known social engineering attacks and a comprehensive overview of advanced social engineering attacks targeting knowledge workers in modern business environments.

Table 2: Top 10 Most Cited Publications (2004-2013)

Ref	Author(s)	Year	Title	Source	Affiliation(s)	Total Citations
[33]	Jagatic, TN; Johnson, NA; Jakobsson, M; Menczer, F	2007	Social phishing	Communications Of The ACM	Massachusetts Institute of Technology (MIT); Indiana University System; Indiana University Bloomington	506
[36]	Anderson, R; Moore, T	2006	The economics of information security	Science	Carnegie Mellon University; Indraprastha Institute of Information Technology Delhi	330
[37]	Xiang, G; Hong, J; Rose, CP; Cranor, L	2011	CANTINA+: A Feature-Rich Machine Learning Framework for Detecting Phishing Web Sites	ACM Transactions On Information and System Security	Carnegie Mellon University	276
[20]	Sheng, S; Holbrook, M; Kumaraguru, P; Cranor, L; Downs, J	2010	Who Falls for Phish? A Demographic Analysis of Phishing Susceptibility and Effectiveness of Interventions	Chi2010: Proceedings Of The 28th Annual Chi Conference On Human Factors In Computing Systems, Vols 1-4	Carnegie Mellon University; Indraprastha Institute of Information Technology Delhi	271
[38]	Grier, C; Thomas, K; Paxson, V; Zhang, M	2010	@spam: The Underground on 140 Characters or Less	Proceedings Of The 17th ACM Conference On Computer And Communications Security (Ccs'10)	University of California System; University of California Berkeley; University of Illinois System; University of Illinois Urbana-Champaign	253
[39]	Bailey, M; Oberheide, J; Andersen, J; Mao, ZM; Jahanian, F; Nazario, J	2007	Automated classification and analysis of Internet malware	Recent Advances In Intrusion Detection, Proceedings	University of Michigan System; University of Michigan	240
[10]	Hong, J	2012	The State of Phishing Attacks	Communications Of The ACM	Carnegie Mellon University; Carnegie Mellon University	222
[40]	Egelman, S; Cranor, LF; Hong, J	2008	You've Been Warned: An Empirical Study of the Effectiveness of Web Browser Phishing Warnings	Chi 2008: 26th Annual Chi Conference On Human Factors In Computing Systems Vols 1 And 2, Conference Proceedings	Carnegie Mellon University	218
[41]	Thomas, K; Grier, C; Ma, J; Paxson, V; Song, D	2011	Design and Evaluation of a Real-Time URL Spam Filtering Service	2011 IEEE Symposium On Security And Privacy (Sp 2011)	University of California System; University of California Berkeley	216
[42]	Vishwanath, A; Herath, T; Chen, R; Wang, JG; Rao, HR	2011	Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model	Decision Support Systems	State University of New York (SUNY) System; State University of New York (SUNY) Buffalo; Brock University; Ball State University; University of Texas System; University of Texas Arlington	212

Table 3: Top 10 Most Cited Publications (2014-2023)

Ref	Author(s)	Year	Title	Source	Affiliation(s)	Total Citations
[35]	Krombholz, K; Hobel, H; Huber, M; Weippl, E	2015	Advanced social engineering attacks	Journal of Information Security And Applications	SBA Research	209
[16]	Sahingoz, OK; Buber, E; Demir, O; Diri, B	2019	Machine learning based phishing detection from URLs	Expert Systems with Applications	Istanbul Kultur University; Marmara University; Yildiz Technical University	191
[43]	Bilge, L; Sen, S; Balzarotti, D; Kirda, E; Kruegel, C	2014	EXPOSURE: A Passive DNS Analysis Service to Detect and Report Malicious Domains	ACM Transactions on Informational and System Security	Hacettepe University; IMT - Institut Mines-Telecom; EURECOM; Northeastern University; University of California System; University of California Santa Barbara	174
[44]	Abdelhamid, N; Ayesh, A; Thabtah, F	2014	Phishing detection based Associative Classification data mining	Expert Systems with Applications	De Montfort University; Canadian University Dubai	162
[45]	Mohammad, RM; Thabtah, F; McCluskey, L	2014	Predicting phishing websites based on self-structuring neural network	Neural Computing & Applications	University of Huddersfield; Canadian University Dubai	140
[46]	Arachchilage, NAG; Love, S	2014	Security awareness of computer users: A phishing threat avoidance perspective	Computers in Human Behavior	University of Oxford; Brunel University	119
[47]	Kumar, BS; Ravi, V	2016	A survey of the applications of text mining in financial domain	Knowledge-Based Systems	University of Hyderabad	119
[48]	Chiew, KL; Tan, CL; Wong, K; Yong, KSC; Tiong, WK	2019	A new hybrid ensemble feature selection framework for machine learning-based phishing detection system	Information Sciences	University of Malaysia Sarawak; Monash University; Monash University Sunway; Curtin University Malaysia	119
[49]	Mandavifar, S; Ghorbani, AA	2019	Application of deep learning to cybersecurity: A survey	Neurocomputing	University of New Brunswick	119
[22]	Alsharnouby, M; Alaca, F; Chiasson, S	2015	Why phishing still works: User strategies for combating phishing attacks	International Journal of Human-Computer Studies	Carleton University	108

4.2 Sources

IEEE Access is the leading journal for the field of phishing, with 83 articles published. Table 4 shows the top ten journals in the field, most of which focus on the technical aspects of security. However, the journal “Computers in Human Behavior” provides a social aspect to this concept.

Table 4: Top 10 Journals

Source	Articles	Citations	CPA
IEEE Access	83	910	11.0
Computers & Security	56	1261	22.5
International Journal Of Advanced Computer Science and Applications	36	137	3.8
International Journal Of Computer Science and Network Security	27	34	1.3
Security and Communication Networks	22	209	9.5
Expert Systems With Applications	21	1002	47.7
Electronics	19	69	3.6
Information and Computer Security	18	116	6.4
Applied Sciences-Basel	17	84	4.9
Computers In Human Behavior	15	579	38.6

CPA: citations per article

Figure 2 is a representation of the relationship of sources based on the number of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

references they have in common. The closer and the stronger the link, the more references they share. The size of the circles represents the total number of citations. Computers & Security has the most citations. Subsequently, Computers & Security, Expert Systems with Applications, and Communications of the ACM have the most references in common.

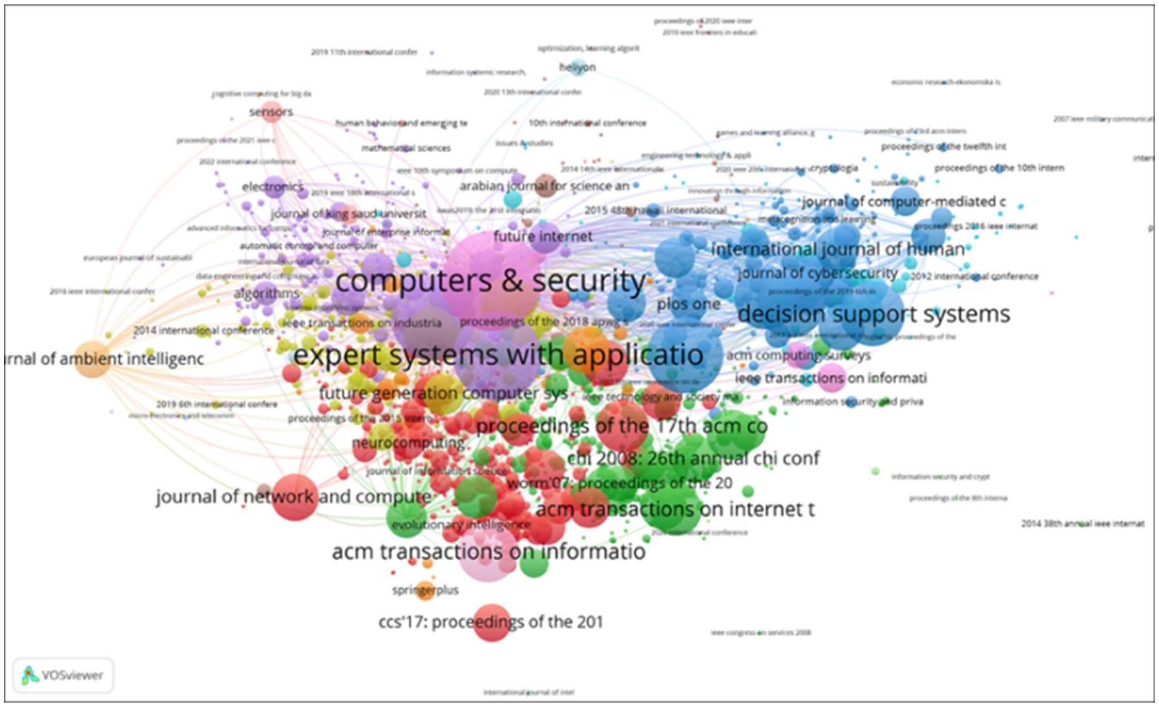
heliyon
sensors
electronics
arabian journal for science an
journal of king saud universit
future internet
mathematical sciences
human behavior and emerging te
10th international conference
journal of computer-mediated c
proceedings of the twelfth int
proceedings of the 10th intern
cryptologia
international journal of human
journal of cybersecurity
2012 international conference
computers & security
algorithms
plos one
decision support systems
ieee transactions on industria
proceedings of the 2018 apwg
2014 international conference
rnal of ambient intelligenc
expert systems with applicatio
acm computing surveys
ieee transactions on informati
information security and priva
future generation computer sys
ieee technology and society ma
proceedings of the 2015 intern
proceedings of the 17th acm co
neurocomputing
chi 2008: 26th annual chi conf
worm'07: proceedings of the 20
journal of network and compute
acm transactions on internet t
evolutionary intelligence
acm transactions on informatio
springerplus
ccs'17: proceedings of the 201
2014 38th annual ieee internat
VOSviewer

Figure 2. Bibliographic coupling of sources

4.3 Authors

Table 5 displays the number of articles and the extent of author collaboration, excluding solo-authored articles, as collaboration can only be assessed with multiple authors involved. The "No collaboration" category indicates articles with authors affiliated with the same institution, whereas "National collaboration" represents articles with authors from different institutions within the same country. Finally, "International collaboration" pertains to articles authored by individuals from distinct countries.

Table 5: Collaboration Type of Articles

Year	No Collaboration	National Collaboration	International Collaboration
2004	0	0	0
2005	13	0	0
2006	22	5	0
2007	38	10	7
2008	44	16	6
2009	43	12	10
2010	39	13	10
2011	35	18	9
2012	52	21	13
2013	64	23	9
2014	79	32	20
2015	103	36	27
2016	107	50	31
2017	122	63	32
2018	127	56	36
2019	166	70	40
2020	148	86	60
2021	168	86	72
2022	159	109	62
2023	35	18	13

Figure 3 graphically represents Table 5, illustrating a noticeable increase in international collaboration between 2019 and 2021, as well as a rise in national collaborations during 2021. Nevertheless, the overall number of internationally collaborative articles remains relatively low compared to articles with authors from the same institution.

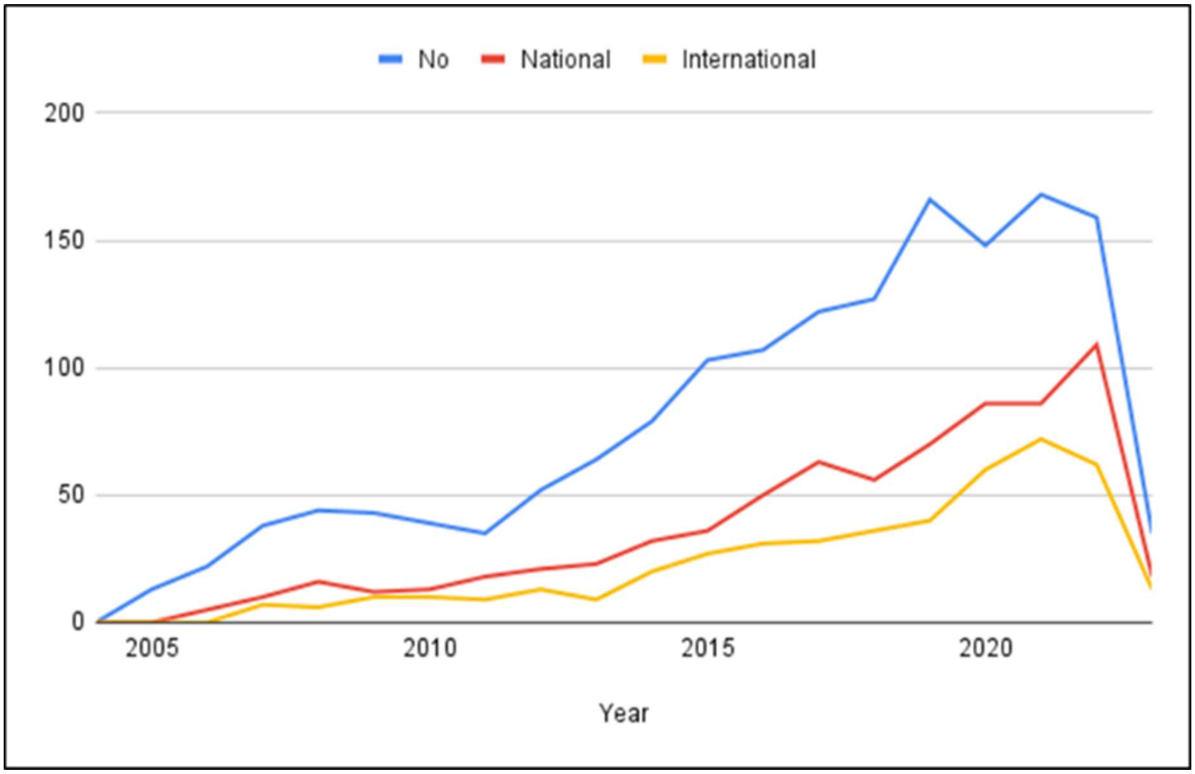

Figure 3. Collaboration type of authors per year

This research employs visualization techniques to differentiate between citation, co-citation, and bibliographic coupling networks. Bibliographic coupling and co-citation are indirect relationships that may offer less accurate insights into article relatedness [50]. A citation link is formed when one item cites another, with VOSviewer treating these links as undirected, meaning no distinction is made between citations from item A to item B or vice versa. Bibliographic coupling links arise when two items cite the same document, while co-citation links emerge when two items are cited by the same document [51].

The visualization of citation relationships among authors shown below reveals Cranor as the most cited author. Each circle symbolizes an author, with larger circles denoting a higher number of citations received. Unique colours represent different clusters, and circle proximity indicates the strength of citation relationships. Additionally, links convey the connections' strength between authors, with more robust lines signifying a higher number of links between the two items. The total link strength of Hong is the highest and the circle can be seen to anchor most authors meaning that this author is very influential and has major connections with many authors in this field.

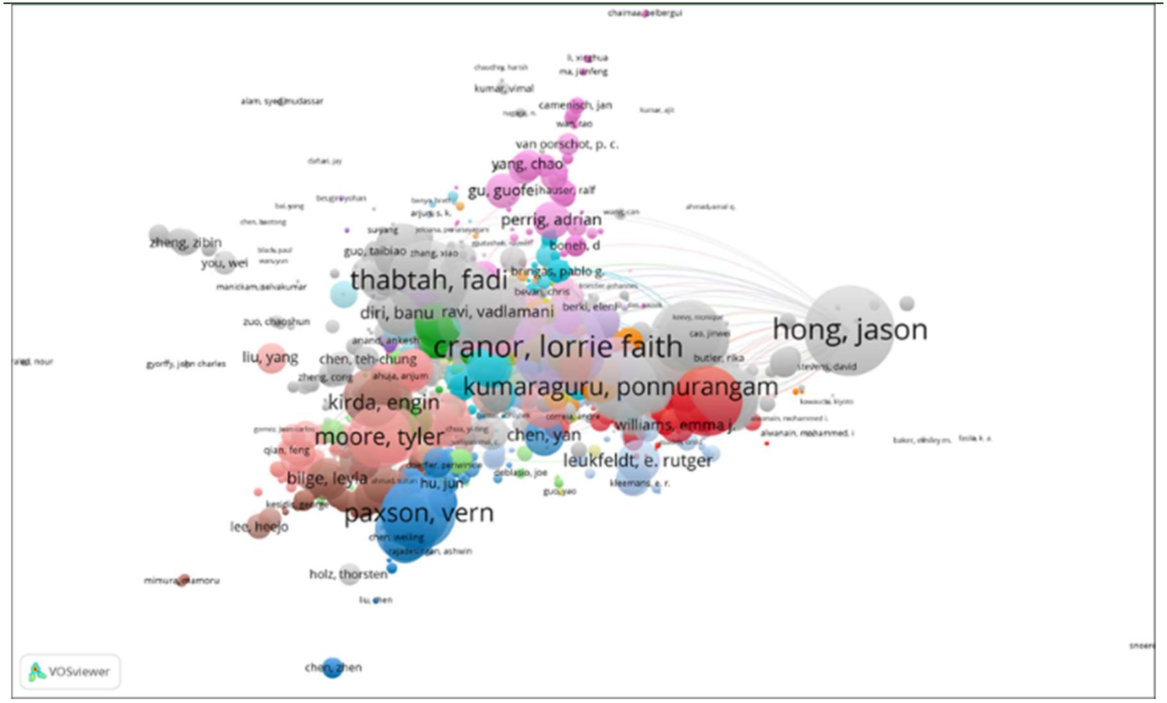

Figure 4. Citation relation of authors

Figure 5's map features circles representing authors, with larger circles signifying a higher number of publications. The closeness of two circles (authors) denotes the strength of their relationship based on bibliographic coupling [52], which means authors closer together in the visualization tend to cite the same publications, while those farther apart usually do not. Bibliographic coupling occurs when two documents both cite a common third document, and this method uses citations to reveal similarities between documents, authors, institutions, or countries. The assumption is that two papers citing the same third paper are closely related and should appear together in a cluster solution on a visualization map. The intensity of bibliographic coupling is measured by the total number of shared references or citations to other third documents [53], and authors in the same cluster typically share similar themes.

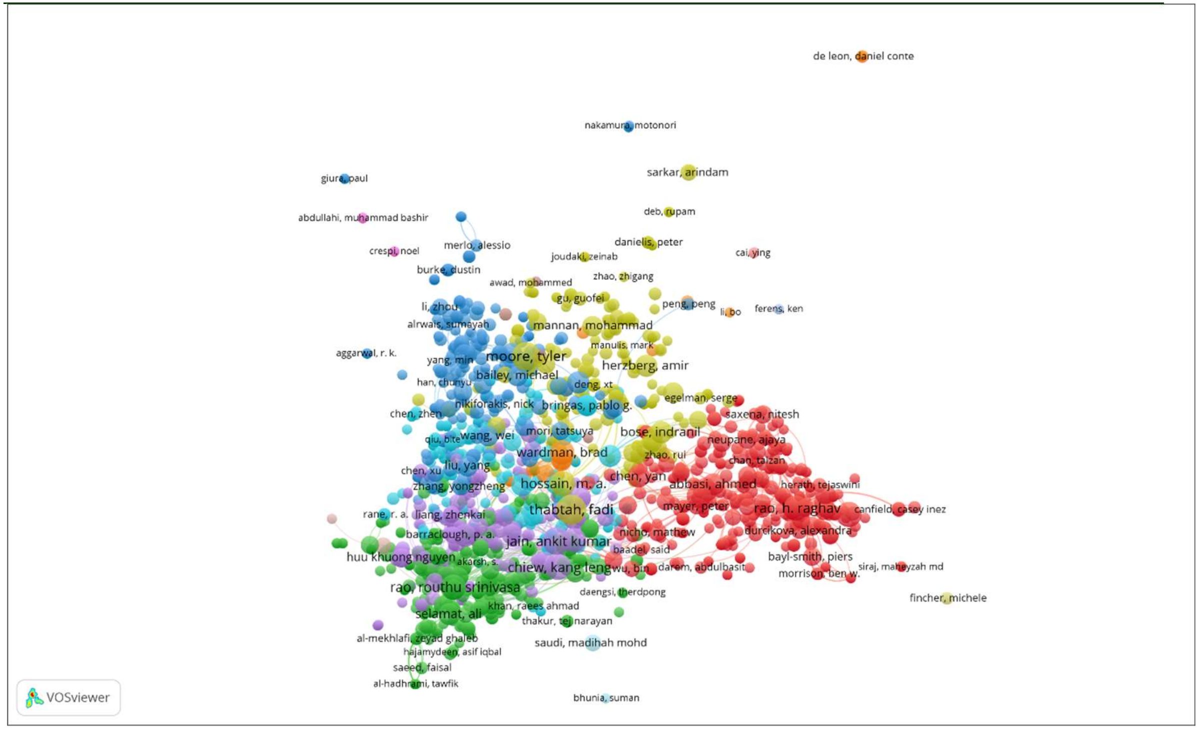

Figure 5. Bibliographic coupling of authors

The co-citation map below illustrates authors who have been most frequently cited together. The closer the authors and the stronger the link, the higher the number of publications that have cited both authors. Co-citation occurs when two documents are cited by a common third document [54], and the analysis of co-citation is based on the assumption that two papers cited together have a strong relationship and should be grouped in a cluster solution on a visualization map. Figure 6 displays a network visualization derived from the co-citation analysis of authors, where each circle or node represents an author and the connections between authors (through co-citations) are depicted by the links between nodes. The proximity between two authors on the map approximately indicates their relatedness in terms of co-citations [55].

Examining the clusters in Figure 6, the green cluster is centered around authors Sheng and Vishwanath, the red cluster around Dhamija, the purple cluster around Mohammed, Chiew, and APWG, and the blue cluster around Zhang and Xiang. It can be inferred that these authors are central to their respective clusters, and other authors within these clusters likely conduct research on similar topics or sub-topics, as they are frequently cited together.

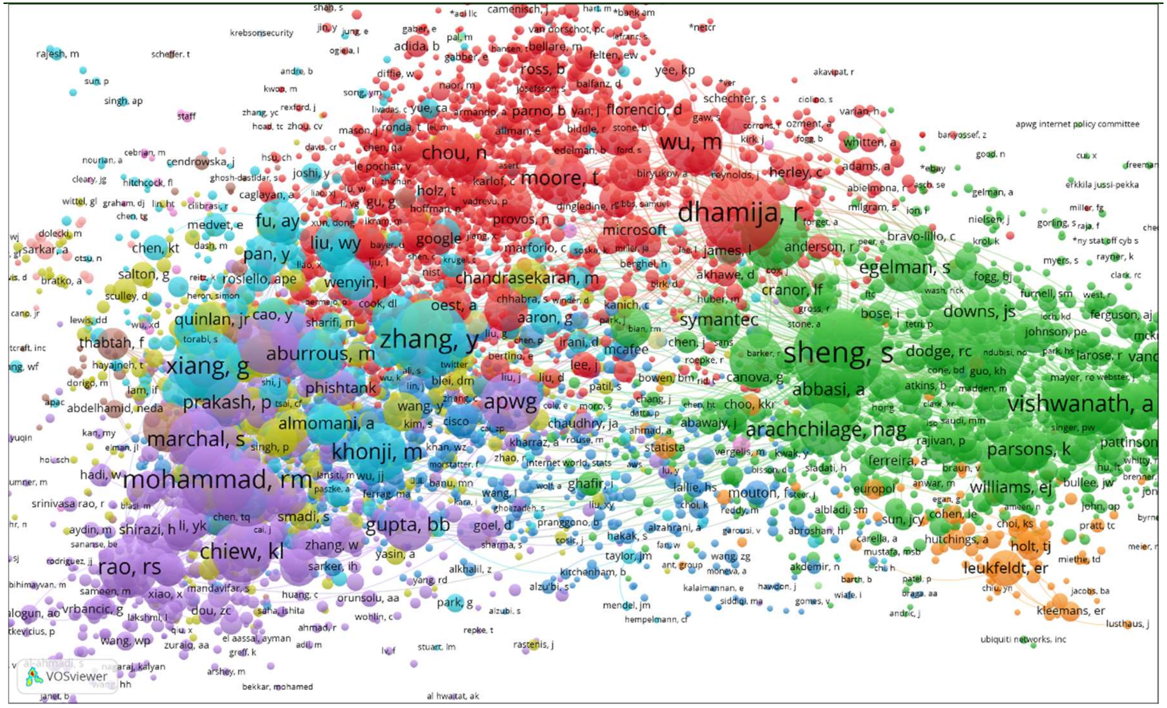

Figure 6. Co-citation analysis of authors

Table 6 lists the top 10 most influential authors in this field based on the total number of citations. Cranor emerges as the most influential author, with 1101 citations from 8 publications. The highest number of publications by a single author is 23. Li is the most productive author in this area of study with respect to number of publications.

Table 6: Top 10 Authors

Author	Articles	Citations	CPA
Cranor, Lf	8	1101	137.6
Hong, J	7	1029	147.0
Thabtah, F	17	822	48.4
Paxson, V	6	686	114.3
Gupta, Bb	16	656	41.0
Kumaraguru, P	9	628	69.8
Vishwanath, A	12	615	51.3
Jakobsson, M	8	611	76.4
Thomas, K	7	591	84.4
Moore, T	15	512	34.1

4.4 Organizations

Analysing the co-authorship map (Figure 7), it is evident that the Chinese Academy of Sciences has collaborated on publications with organizations from various clusters, such as the University of Purdue and University of Ilorin.

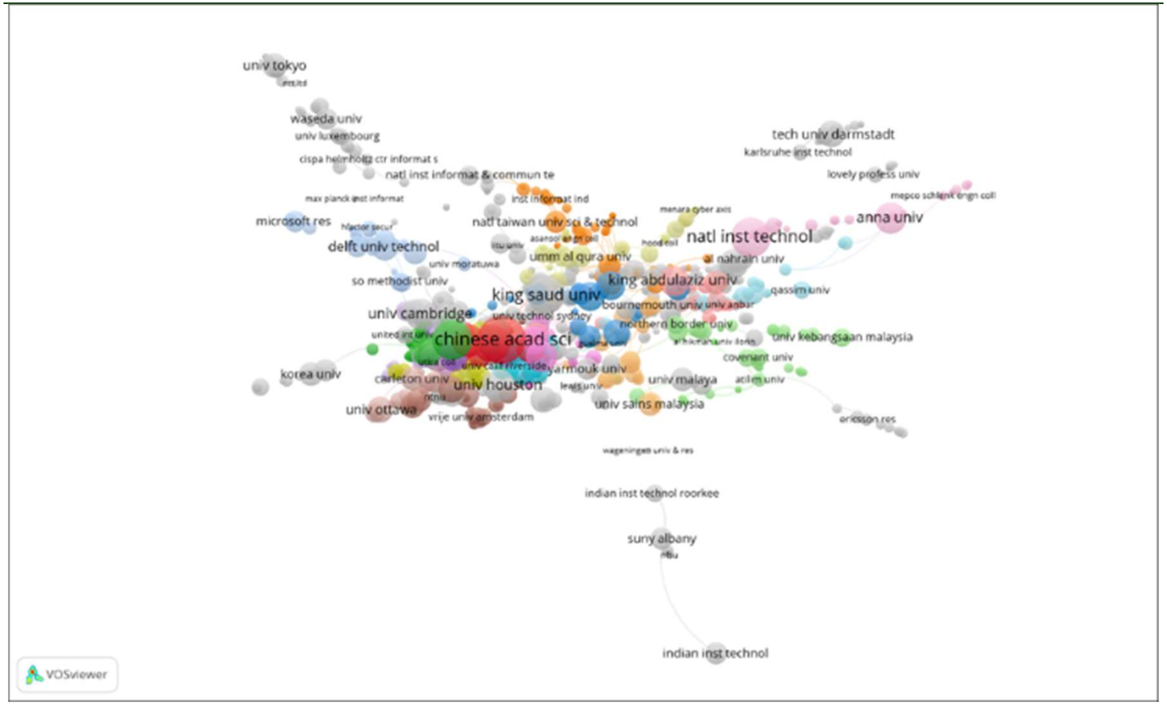

Figure 7. Co-authorship analysis of organizations

Figure 8 reveals that Carnegie Mellon University shares a significant number of references with other organizations, as it serves as the anchor of the network.

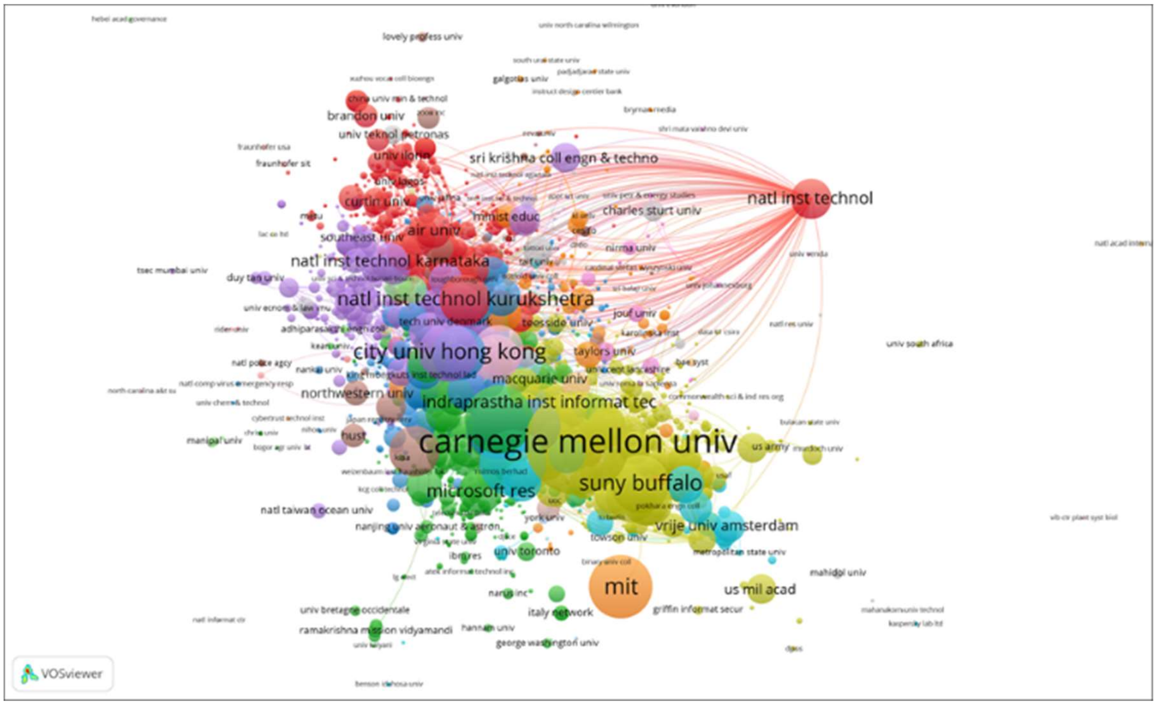

Figure 8. Bibliographic coupling analysis of organizations

Carnegie Mellon University holds the most citations and published articles on this subject (Table 7).

Table 7: Top 30 Organizations

Organization	Country	Articles	Citations	CPA
Carnegie Mellon University	USA	31	1806	58.3
Indiana University	USA	22	778	35.4
University Of Cambridge	England	19	745	39.2
MIT	USA	8	674	84.3
University California Berkeley	USA	10	661	66.1
City University Hong Kong	Hong Kong	15	580	38.7
Suny Buffalo	USA	12	569	47.4
University Of Illinois	USA	15	539	35.9
University Of Michigan	USA	15	489	32.6
University Of Malaya	Malaysia	11	480	43.6
University Of California Santa Barbara	USA	6	432	72.0
University Of Huddersfield	England	9	428	47.6
National Institute Of Technology, Kurukshetra	India	10	426	42.6
University Of Texas Arlington	USA	12	423	35.3
University Of Malaysia Sarawak	Malaysia	14	399	28.5
Canadian University Dubai	UAE	7	374	53.4
King Saud University	Saudi Arabia	28	362	12.9
International Computer Science Institute	USA	8	335	41.9
Brock University	Canada	5	331	66.2
Ball State University	USA	4	329	82.3
Arbor Networks	USA	2	323	161.5
Microsoft Research	USA	9	319	35.4
Purdue University	USA	21	319	15.2
Chinese Academy Of Sciences	China	45	295	6.6
Indraprastha Institute Of Information Technology	India	3	289	96.3
National Institutes Of Technology	India	31	283	9.1
Harvard University	USA	6	264	44.0
Deakin University	Australia	17	255	15.0
Vienna University Of Technology	Austria	7	254	36.3
University Of Virginia	USA	13	253	19.5

4.5 Countries

The co-authorship map displays the most collaborative countries in this field, with the size of the nodes representing the number of publications and the node colour indicating different clusters. Countries within the same cluster can be considered to have collaborated more frequently on publications. The USA is the most collaborative country, with a total of 48 links and the strongest collaboration with China and India.

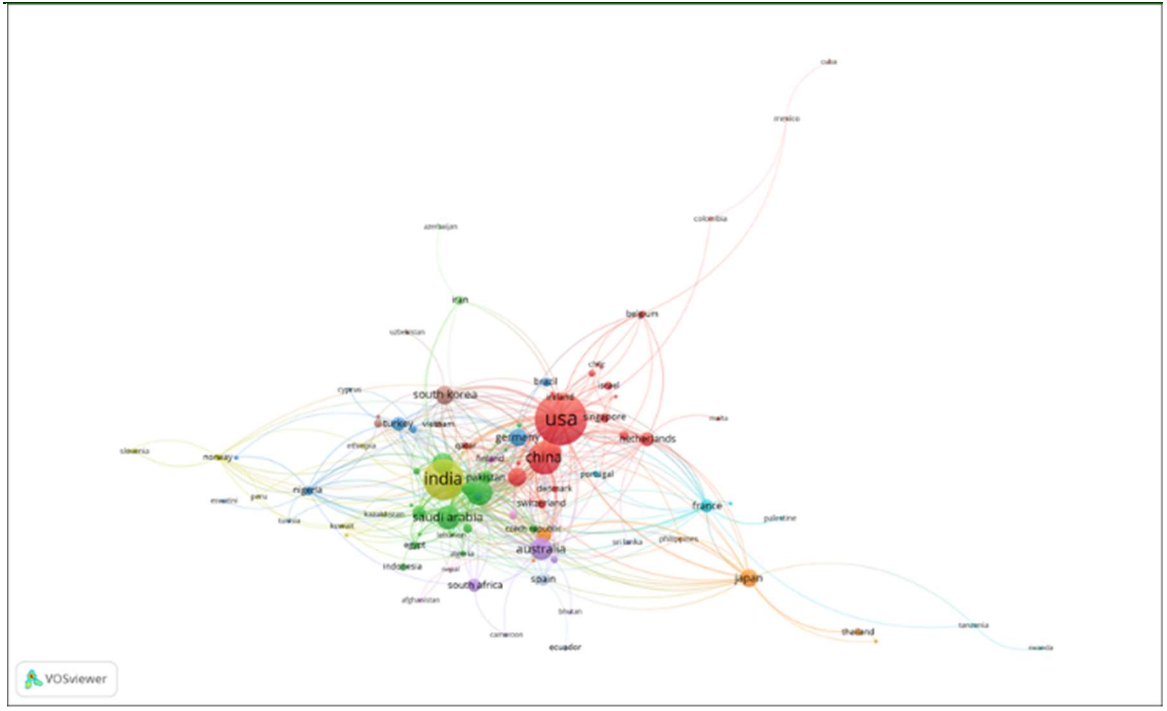

Figure 9. Co-authorship analysis of countries

The USA is also the most prolific country, amassing 12924 citations across 787 articles. Table 8 lists the top 30 countries in terms of the total number of citations.

Table 8: Top 30 Countries

Country	**Articles**	**Citations**	**CPA**
USA	787	12924	16.4
England	196	3769	19.2
India	478	3711	7.8
China	321	3291	10.3
Australia	141	1688	12.0
Malaysia	123	1579	12.8
Canada	103	1502	14.6
Saudi Arabia	162	1170	7.2
South Korea	102	1013	9.9
Taiwan	63	801	12.7
UAE	49	794	16.2
Netherlands	57	732	12.8
Jordan	57	731	12.8
France	48	639	13.3
Austria	27	636	23.6
Turkey	55	625	11.4
Germany	90	575	6.4
Italy	55	517	9.4
Pakistan	49	492	10.0
Spain	46	383	8.3
South Africa	52	377	7.3
Iran	29	350	12.1
Singapore	28	304	10.9
Switzerland	26	275	10.6

4.6 Keywords

Identifying emerging research fronts is crucial for understanding research efforts within a specific scientific domain [25]. Figure 10 showcases a map of co-occurring keywords. The keyword "phishing" was excluded from the list as it was the primary search term, and as expected, would be the most common keyword among the articles. Furthermore, keywords were checked for spelling differences and variations of words were grouped together to form a single version. Finally, the subtopics of keywords were consolidated into single overarching keywords. The resulting taxonomy incorporates these combined subtopics as part of the broader terms. Table 9 displays the top 20 keywords, with "security" being the most frequently occurring term. This is unsurprising, considering that phishing is a significant aspect of security research.

Table 9: Top 20 Keywords

keyword	occurrences
Security	646
Detection	411
Machine Learning	388
Attack	355
Cybersecurity	293
Feature	213
Classification	212
Web	180
Social Engineering	174
Cyber	151
Model	149
Algorithm	146
Neural Network	143
Authentication	141
Email	138
Malware	132
Spam	131
Behavio(u)r	114
Anti-Phishing	108
URL	108

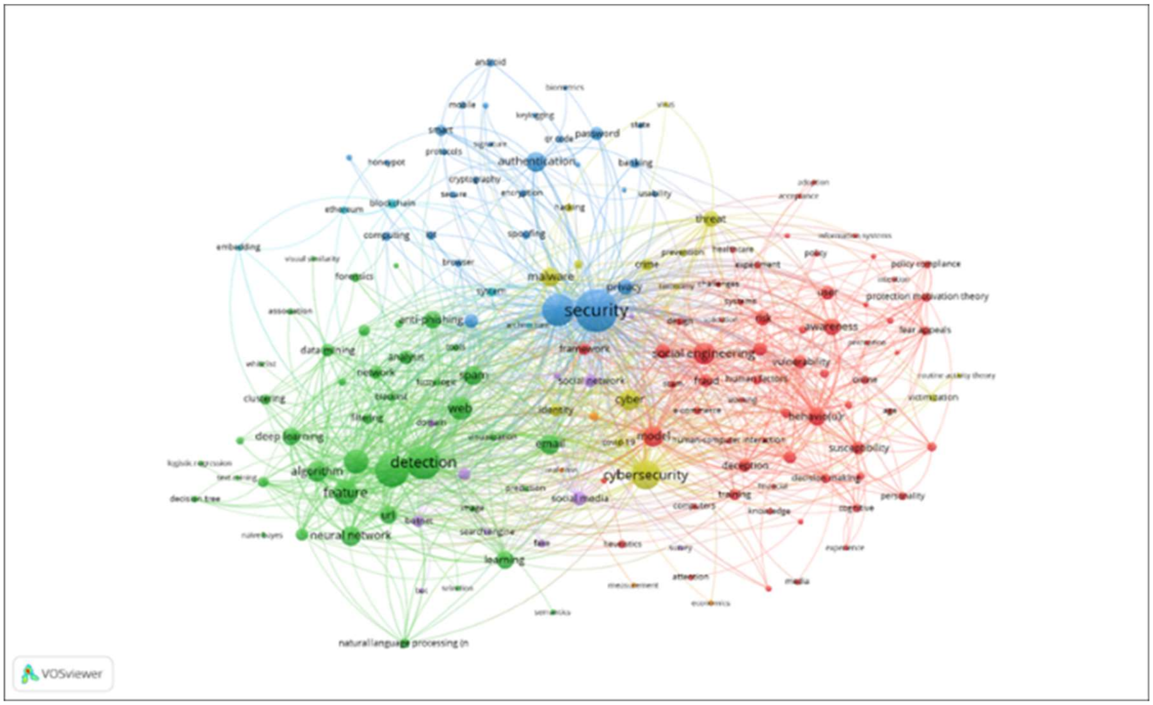

Figure 10. Co-occurrence analysis of keywords

The Vosviewer software divided the keywords into six clusters, which can be found in Table 10 along with their corresponding keywords.

Table 10: Keyword Clusters

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Keywords	acceptance, accuracy, adoption, age, attention, attitude, awareness, behavio(u)r, challenges, cognitive, computers, countermeasures, culture, deception, decision making, design, deterrence, e-commerce, education, experience, experiment, fear appeals, financial, framework, fraud, healthcare, heuristics, human factors, human-computer interaction, impact, individual-differences, information, information systems, intention, internet, knowledge, media, model, online, perception, performance, personality, persuasion, performance, personality, persuasion, policy, policy compliance, protection motivation theory, psychology, review, risk, scam, science, self-efficacy, social engineering, susceptibility, systems, technology, training, trust, user, validation, vulnerability, warning	algorithm, analysis, anti-phishing, artificial intelligence, association, big data, blacklist, classification, classifier, clustering, data mining, decision tree, deep learning, detection, email, feature, filtering, forensics, fuzzy logic, identification, image, intrusion detection, learning, logistic regression, machine learning, naive bayes, natural language processing, network, neural network, optimization, prediction, selection, semantics, spam, support vector machine, text mining, tools, url, visual similarity, visualization, web, whitelist	android, attack, authentication, banking, biometrics, browser, captcha, certificate, computing, cryptography, encryption, honeypot, iot, keylogging, malicious, mobile, password, privacy, protocols, qr code, scheme, secure, security, signature, smart, spoofing, state, usability, visual cryptography	covid-19, crime, cyber, cybersecurity, hacking, identity, malware, prevention, ransomware, routine activity theory, taxonomy, threat, victimization, virus	bot, botnet, dataset, dns, domain, fake, issues, search engine, social media, social network, survey	architecture, blockchain, embedding, ethereum, system

Cluster 1 – "Human Factors and User Behavior": This cluster focuses on human factors, user behaviour, and psychological aspects in relation to cybersecurity, technology, and online activities. The themes include user awareness, behaviour, education, decision-making, risk perception, social engineering, and policy compliance.

Cluster 2 – "Detection and Prevention Techniques": This cluster is centred around algorithms, machine learning techniques, and data analysis methods used in cybersecurity for detecting and preventing phishing attacks, spam, and other threats. The themes include classification, clustering, deep learning, feature selection, and natural language processing.

Cluster 3 – "Security Measures and Authentication": This cluster emphasizes security measures, authentication methods, and technologies used to protect users, data, and systems from cyber threats. The themes include encryption, biometrics, cryptography, IoT security, and secure protocols.

Cluster 4 – "Cyber Threats and Crime": This cluster deals with various types of cyber threats, cybercrimes, and their consequences. The themes include hacking, identity theft, malware, ransomware, and victimization.

Cluster 5 – "Online Platforms and Social Media": This cluster is related to online platforms, social media, search engines, and the challenges and issues arising from their use. The themes include bots, botnets, fake news, social networks, and domain spoofing.

Cluster 6 – "Emerging Technologies and Systems": This cluster focuses on emerging technologies, systems, and architectures used in cybersecurity and related fields. The themes include blockchain, Ethereum, and system embedding.

Based on the analysed articles, the following taxonomy has been developed for phishing research (Figure 11). There are a total of seven categories: human factors, attack vectors, detection and prevention, theories, emerging technologies and computing, targets, and review and research methodologies. This taxonomy provides a comprehensive insight into the topics and subtopics investigated within the phishing literature. It offers an understanding of the theoretical frameworks employed and the machine learning and data mining techniques applied, as well as the human elements involved in phishing.

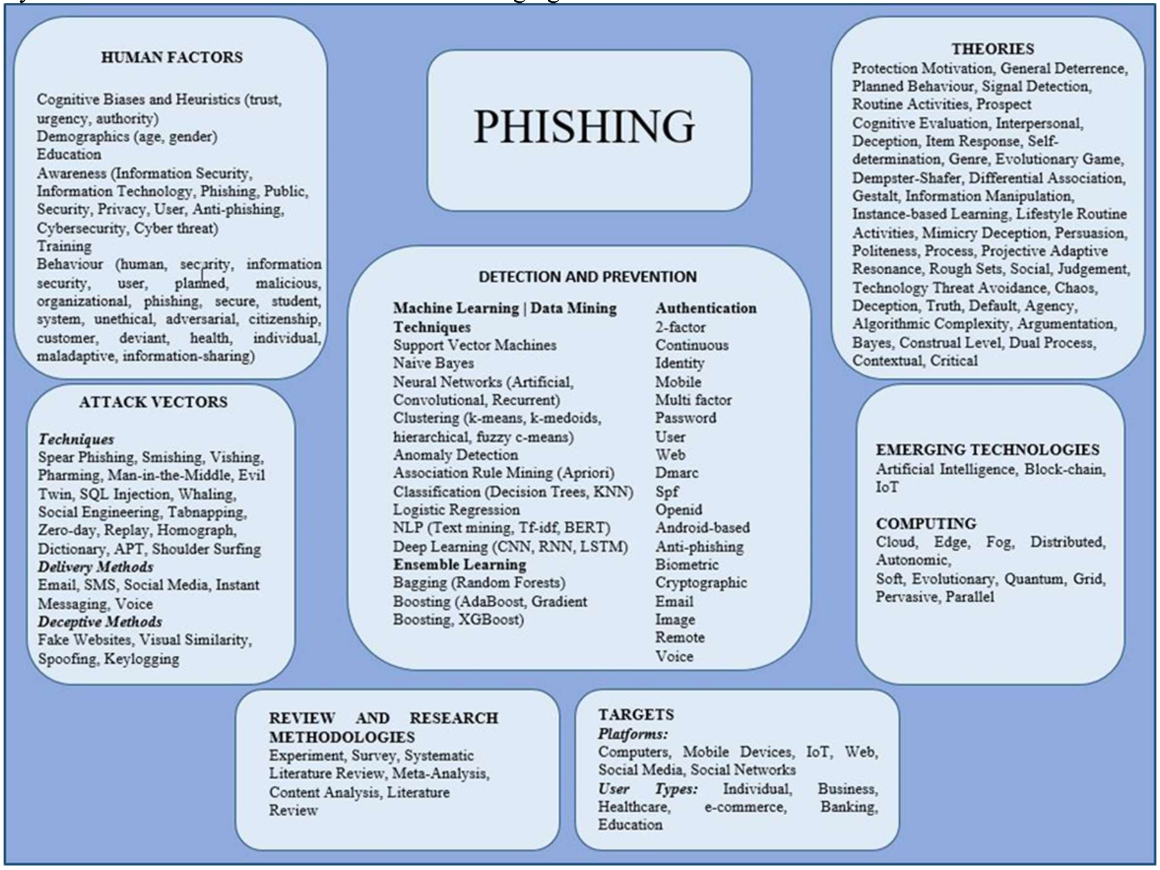
PHISHING
HUMAN FACTORS
Cognitive Biases and Heuristics (trust, urgency, authority)
Demographics (age, gender)
Education
Awareness (Information Security, Information Technology, Phishing, Public, Security, Privacy, User, Anti-phishing, Cybersecurity, Cyber threat)
Training
Behaviour (human, security, information security, user, planned, malicious, organizational, phishing, secure, student, system, unethical, adversarial, citizenship, customer, deviant, health, individual, maladaptive, information-sharing)
ATTACK VECTORS
Techniques
Spear Phishing, Smishing, Vishing, Pharming, Man-in-the-Middle, Evil Twin, SQL Injection, Whaling, Social Engineering, Tabnapping, Zero-day, Replay, Homograph, Dictionary, APT, Shoulder Surfing
Delivery Methods
Email, SMS, Social Media, Instant Messaging, Voice
Deceptive Methods
Fake Websites, Visual Similarity, Spoofing, Keylogging
DETECTION AND PREVENTION
Machine Learning | Data Mining Techniques
Support Vector Machines
Naive Bayes
Neural Networks (Artificial, Convolutional, Recurrent)
Clustering (k-means, k-medoids, hierarchical, fuzzy c-means)
Anomaly Detection
Association Rule Mining (Apriori)
Classification (Decision Trees, KNN)
Logistic Regression
NLP (Text mining, Tf-idf, BERT)
Deep Learning (CNN, RNN, LSTM)
Ensemble Learning
Bagging (Random Forests)
Boosting (AdaBoost, Gradient Boosting, XGBoost)
Authentication
2-factor
Continuous
Identity
Mobile
Multi factor
Password
User
Web
Dmarc
Spf
Openid
Android-based
Anti-phishing
Biometric
Cryptographic
Email
Image
Remote
Voice
THEORIES
Protection Motivation, General Deterrence, Planned Behaviour, Signal Detection, Routine Activities, Prospect
Cognitive Evaluation, Interpersonal, Deception, Item Response, Self-determination, Genre, Evolutionary Game, Dempster-Shafer, Differential Association, Gestalt, Information Manipulation, Instance-based Learning, Lifestyle Routine Activities, Mimicry Deception, Persuasion, Politeness, Process, Projective Adaptive Resonance, Rough Sets, Social, Judgement, Technology Threat Avoidance, Chaos, Deception, Truth, Default, Agency, Algorithmic Complexity, Argumentation, Bayes, Construal Level, Dual Process, Contextual, Critical
EMERGING TECHNOLOGIES
Artificial Intelligence, Block-chain, IoT
COMPUTING
Cloud, Edge, Fog, Distributed, Autonomic, Soft, Evolutionary, Quantum, Grid, Pervasive, Parallel
REVIEW AND RESEARCH METHODOLOGIES
Experiment, Survey, Systematic Literature Review, Meta-Analysis, Content Analysis, Literature Review
TARGETS
Platforms:
Computers, Mobile Devices, IoT, Web, Social Media, Social Networks
User Types: Individual, Business, Healthcare, e-commerce, Banking, Education

Figure 11. Taxonomy of phishing research

5. DISCUSSION AND CONCLUSION

Phishing attacks are a pervasive problem demanding a holistic, international approach to address them. Investigating global academic studies on such incidents helps us comprehend the magnitude of this issue across the world, while also considering the distinct challenges and viewpoints present in various regions and universities.

In this bibliometric study, various analyses were conducted to understand the trends, themes, and patterns in phishing research from 2004 to 2023. The study aimed to explore the growth and development of the research domain, identify influential authors and institutions, examine collaboration patterns, and detect emerging research fronts.

The growth and trends in the phishing research domain were demonstrated by analysing the number of articles published and their citations. A consistent upward trajectory in the number of articles signifies the increased interest and attention from researchers and the academic community. However, the total citations showed some fluctuations, which could be due to various factors such as the time lag between publication and citation, interdisciplinary nature of the papers, or the emergence of new research areas.

To further understand the research landscape, the most cited articles were divided into two distinct 10-year intervals. The analysis of the most cited articles from these intervals provided insights into the key research topics and the impact of these articles on the field of phishing. It is worth noting that the most cited article during the first period [33] sought to ethically determine the extent to which social context could enhance the success of a phishing attack, focusing on Indiana University students aged 18 to 24 years old. The results highlighted the necessity for wide-ranging educational campaigns about phishing and other security threats to increase awareness and minimize vulnerabilities among internet users. On the other hand, the most cited article in the second period [35] examined the escalating menace of social engineering in virtual communities, particularly as a result of the growing prevalence of BYOD policies and online communication tools. This study provided a classification of well-known social engineering attacks and a thorough overview of sophisticated social engineering attacks aimed at knowledge workers in contemporary business settings.

Journal analysis revealed that the majority of publications in the field of phishing are published in technical security-focused journals. The leading journal in this area is IEEE Access. The analysis of the relationship of sources based on the number of references they have in common showed that Computers & Security, Expert Systems with Applications, and Communications of the ACM shared the most references. Collaboration patterns were examined by analysing the extent of author collaboration in terms of national and international partnerships. A noticeable increase in international collaboration was observed between 2019 and 2021. The citation relationships among authors were visualized, revealing the most cited and influential authors in the field, the most prolific being Cranor and Li. The co-authorship map displayed the most collaborative institutions and countries in this field, with the Chinese Academy of Sciences and the USA being the most collaborative institution and country, respectively.

The observation that most publications in the phishing research domain come from universities, with only a few non-university organizations among the top 30, suggests that there may be untapped potential for greater collaboration between academia and industry. In addition, the involvement of governmental and non-governmental organizations in phishing research could be explored further.

Emerging research fronts were identified by analysing co-occurring keywords in the articles, providing a comprehensive taxonomy of phishing research. The taxonomy consisted of seven categories, namely human factors and user behaviour, detection and prevention techniques, security measures and authentication, cyber threats and crime, online platforms and social media, emerging technologies and systems, and economics and protection strategies.

In summary, this bibliometric study provides valuable insights into the phishing research domain, highlighting its growth, influential authors and institutions, collaboration patterns, and emerging research fronts. The results of the study can serve as a basis for future research directions, collaboration opportunities, and strategic decision-making in the field of phishing.

The results, findings, and future research directions from the bibliometric study have both research and practical implications that can contribute to the ongoing efforts in combating phishing attacks.

5.1 Research Implications

Expanding interdisciplinary collaboration: The analysis demonstrated a surge in both international and national collaborations in the last few years; however, the overall number of internationally collaborative articles remained low in comparison to publications with authors all originating from the same institution. Researchers should consider working across disciplines to develop holistic approaches to address phishing challenges. By combining technical expertise with insights from psychology, sociology, and economics, researchers can create more effective strategies that consider various aspects of phishing attacks and human behaviour.

Encouraging international partnerships: To better understand and combat phishing threats on a global scale, researchers should foster international collaboration. Collaborating across borders allows researchers to share knowledge, expertise, and resources, resulting in more comprehensive and diverse solutions to phishing challenges.

Emphasizing emerging research fronts: Researchers should focus on the emerging research fronts identified in the study, such as human factors, detection and prevention techniques, security measures, and authentication methods. By concentrating on these areas, researchers can address the most pressing issues and advance the state of the art in phishing research.

Incorporating emerging technologies: The adoption of emerging technologies, such as blockchain and IoT, in phishing research can lead to innovative solutions and improved cybersecurity. Researchers should explore the potential applications of these technologies in the context of phishing to stay ahead of evolving threats.

Promoting academia-industry-government partnerships: To foster innovation and accelerate the practical application of phishing research, future studies may explore the possibility of creating stronger partnerships among academic institutions, industry players, and governmental/non-governmental organizations. These collaborations can facilitate the exchange of knowledge, resources, and expertise, bridging the gap between theoretical research and real-world challenges.

Global research & regional insights on phishing attacks: Phishing attacks represent a widespread issue that requires a comprehensive, global strategy to tackle them. Examining international academic research on these occurrences allows us to understand the extent of the problem worldwide and take into account the unique obstacles and perspectives found in different regions and academic institutions.

5.2 Practical Implications

Developing educational programs: The findings on human factors and user behaviour suggest that there is a need for effective educational programs to raise awareness about phishing threats. Organizations, educational institutions, and governments can develop and implement training programs to help individuals recognize and avoid phishing attacks.

Enhancing phishing detection systems: Organizations should invest in the development and implementation of advanced phishing detection and prevention techniques. By adopting cutting-edge algorithms, machine learning models, and data analysis methods, organizations can better protect their users, data, and systems from phishing attacks.

Implementing robust security measures: Businesses and organizations should prioritize the implementation of robust security measures and authentication methods to safeguard against phishing attacks. By employing multi-factor authentication, secure protocols, and encryption technologies, organizations can reduce the likelihood of successful phishing attacks.

Strengthening cybersecurity policies: Policymakers and regulatory bodies should consider the findings of this study when developing and updating cybersecurity policies. By incorporating insights from emerging research fronts and future research directions, policies can become more comprehensive and effective at addressing the challenges posed by phishing attacks.

By considering these research and practical implications, stakeholders in the phishing research community, organizations, and policymakers can work together to create more effective strategies and solutions to address the ever-evolving landscape of phishing threats.

In light of the objectives, this bibliometric study has provided an encompassing view of the phishing research landscape. It revealed the domain's growth, identified principal authors and institutions, highlighted collaboration trends, and spotlighted emerging areas of interest.

However, despite its breadth, the study underscores the need for:

A more integrated approach between academia and industry for a fuller comprehension of the phishing phenomenon.

A diversified data source approach in future bibliometric studies to capture a richer set of insights.

An emphasis on human-centered and technology-driven solutions for phishing, especially considering the nuanced challenges posed by evolving online threats.

To conclude, while this study offers a solid foundation, it also serves as a stepping stone and an invitation for more in-depth, collaborative, and multidisciplinary research in the phishing realm. The continual evolution of phishing strategies necessitates an agile, informed, and collaborative response from the global research community.

6. LIMITATIONS

This bibliometric analysis has certain limitations that warrant consideration when interpreting its outcomes. The study relies exclusively on the Web of Science (WoS) database; including other databases, such as Scopus or Google Scholar, could have presented a broader perspective on phishing literature. The network analysis methods employed in this investigation, including co-citation, co-authorship, co-occurrence, and bibliographic coupling, are dependent on quantitative data and might not fully reflect the qualitative aspects or subtle nuances of the research area. Additionally, WoS data only accounts for the first author of a cited document, limiting the co-citation analysis to consider only the first author. The study also depends on citation data, which can be influenced by various factors such as self-citations, journal impact factors, or the popularity of specific research topics, rather than the genuine quality or relevance of the research itself.

The consistency in the upward trend of phishing-related articles suggests the field's growth, but citation fluctuations indicate that not all research contributions gain equal recognition or relevance.

Relying solely on two 10-year intervals to analyse the most cited articles might have introduced a myopic view. Some influential papers from mid-decades might have been overlooked.

The field seems to have a predominant academic focus, as indicated by most publications originating from universities. However, the minimal involvement of non-university entities, including industry partners, presents a potential shortcoming in our understanding of practical challenges and solutions related to phishing.

Our focus on Web of Science as the sole database might have introduced a selection bias, potentially excluding other significant works and trends present in databases like Scopus or Google Scholar.

REFERENCES:

[1] Ahmetoglu H, Das R. A comprehensive review on detection of cyber-attacks: Data sets, methods, challenges, and future research directions. Internet of Things. 2022 Sep 20:100615.

[2] APWG. Phishing Activity Trends Report, 3rd Quarter 2022. Lexington: APWG; 2022.

[3] Almomani A, Alauthman M, Shatnawi MT, Alweshah M, Alrosan A, Alomoush W, Gupta BB. Phishing website detection with semantic features based on machine learning classifiers: A comparative study. International Journal on Semantic Web and Information Systems (IJSWIS). 2022 Jan 1;18(1):1-24.

[4] Pattinson M, Jerram C, Parsons K, McCormac A, Butavicius M. Why do some people manage phishing e-mails better than others?. Information Management & Computer Security. 2012 Mar 16;20(1):18-28.

[5] Halevi T, Lewis J, Memon N. A pilot study of cyber security and privacy related behavior and personality traits. InProceedings of the 22nd international conference on world wide web 2013 May 13 (pp. 737-744).

[6] Iuga C, Nurse JR, Erola A. Baiting the hook: factors impacting susceptibility to phishing attacks. Human-centric Computing and Information Sciences. 2016 Dec;6:1-20.

[7] Jampen D, Gür G, Sutter T, Tellenbach B. Don't click: towards an effective anti-phishing training. A comparative literature review. Human-centric Computing and Information Sciences. 2020 Dec;10(1):1-41.

[8] Zupic I, Čater T. Bibliometric methods in management and organization. Organizational research methods. 2015 Jul;18(3):429-72.

[9] Pejić-Bach M, Jajić I, Kamenjarska T. A Bibliometric Analysis of Phishing in the Big Data Era: High Focus on Algorithms and Low Focus on People. Procedia Computer Science. 2023 Jan 1;219:91-8.

[10] Hong J. The state of phishing attacks. Communications of the ACM. 2012 Jan 1;55(1):74-81.

[11] Pienta D, Thatcher JB, Johnston AC. A taxonomy of phishing: Attack types spanning economic, temporal, breadth, and target boundaries.

[12] Chiew KL, Yong KS, Tan CL. A survey of phishing attacks: Their types, vectors and technical approaches. Expert Systems with Applications. 2018 Sep 15;106:1-20.

[13] Aleroud A, Zhou L. Phishing environments, techniques, and countermeasures: A survey. Computers & Security. 2017 Jul 1;68:160-96.

[14] Gupta BB, Arachchilage NA, Psannis KE. Defending against phishing attacks: taxonomy of methods, current issues and future directions. Telecommunication Systems. 2018 Feb;67:247-67.

[15] Abu-Nimeh S, Nappa D, Wang X, Nair S. A comparison of machine learning techniques for phishing detection. InProceedings of the anti-phishing working groups 2nd annual eCrime researchers summit 2007 Oct 4 (pp. 60-69).

[16] Sahingoz OK, Buber E, Demir O, Diri B. Machine learning based phishing detection from URLs. Expert Systems with Applications. 2019 Mar 1;117:345-57.

[17] Shahrivari V, Darabi MM, Izadi M. Phishing detection using machine learning techniques. arXiv preprint arXiv:2009.11116. 2020 Sep 20.

[18] Paliath S, Qbeitah MA, Aldwairi M. PhishOut: Effective phishing detection using selected features. In2020 27th International Conference on Telecommunications (ICT) 2020 Oct 5 (pp. 1-5). IEEE.

[19] Sheng S, Magnien B, Kumaraguru P, Acquisti A, Cranor LF, Hong J, Nunge E. Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. InProceedings of the 3rd symposium on Usable privacy and security 2007 Jul 18 (pp. 88-99).

[20] Sheng S, Holbrook M, Kumaraguru P, Cranor LF, Downs J. Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. InProceedings of the SIGCHI conference on human factors in computing systems 2010 Apr 10 (pp. 373-382).

[21] Halevi T, Lewis J, Memon N. Phishing, personality traits and Facebook. arXiv preprint arXiv:1301.7643. 2013 Jan 31.

[22] Alsharnouby M, Alaca F, Chiasson S. Why phishing still works: User strategies for combating phishing attacks. International Journal of Human-Computer Studies. 2015 Oct 1;82:69-82.

[23] Mahmud M, Haq IU. Information security in business: a bibliometric analysis of the 100 top cited articles. Library Philosophy and Practice. 2021:1-49.

[24] Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research. 2021 Sep 1;133:285-96.

[25] Munshi A, Singla AR, Trivedi KJ, Jegede OO, Abodunde OO, Sonkar SK, Kumar S, Mahala A, Tripathi M, Ramkumar S, Rahimi S. Scientometric-based Knowledge Map of Food Science and Technology Research in India. Journal of Scientometric Research. 2022 Sep;11(3):409-18.

[26] Baker HK, Kumar S, Pandey N. Forty years of the journal of futures markets: a bibliometric overview. Journal of Futures Markets. 2021 Jul;41(7):1027-54.

[27] Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F. Science mapping software tools: Review, analysis, and cooperative study among tools. Journal of the American Society for information Science and Technology. 2011 Jul;62(7):1382-402.

[28] Pranckutė R. Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. Publications. 2021 Mar 12;9(1):12.

[29] Foster AT. Artificial intelligence in project management. Cost Engineering. 1988 Jun 1;30(6):21.

[30] Van Eck NJ, Waltman L. VOSviewer manual. Leiden: Univeristeit Leiden. 2013 Nov;1(1):1-53.

[31] Healy M, Hammer S, McIlveen P. Mapping graduate employability and career development in higher education research: A citation network analysis. Studies in Higher Education. 2022 Apr 3;47(4):799-811.

[32] Van Noorden R. Interdisciplinary research by the numbers. Nature. 2015 Sep 17;525(7569):306-7.

[33] Jagatic TN, Johnson NA, Jakobsson M, Menczer F. Social phishing. Communications of the ACM. 2007 Oct 1;50(10):94-100.

[34] Chawla S, Mehrotra M. A comprehensive science mapping analysis of textual emotion

mining in online social networks. Int. J. Adv. Comput. Sci. Appl. 2020;11(5):218-29.

[35] Krombholz K, Hobel H, Huber M, Weippl E. Advanced social engineering attacks. Journal of Information Security and applications. 2015 Jun 1;22:113-22.

[36] Anderson R, Moore T. The economics of information security. science. 2006 Oct 27;314(5799):610-3.

[37] Xiang G, Hong J, Rose CP, Cranor L. Cantina+ a feature-rich machine learning framework for detecting phishing web sites. ACM Transactions on Information and System Security (TISSEC). 2011 Sep 1;14(2):1-28.

[38] Grier C, Thomas K, Paxson V, Zhang M. @ spam: the underground on 140 characters or less. InProceedings of the 17th ACM conference on Computer and communications security 2010 Oct 4 (pp. 27-37).

[39] Bailey M, Oberheide J, Andersen J, Mao ZM, Jahanian F, Nazario J. Automated classification and analysis of internet malware. InRecent Advances in Intrusion Detection: 10th International Symposium, RAID 2007, Gold Goast, Australia, September 5-7, 2007. Proceedings 10 2007 (pp. 178-197). Springer Berlin Heidelberg.

[40] Egelman S, Cranor LF, Hong J. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. InProceedings of the SIGCHI Conference on Human Factors in Computing Systems 2008 Apr 6 (pp. 1065-1074).

[41] Thomas K, Grier C, Ma J, Paxson V, Song D. Design and evaluation of a real-time url spam filtering service. In2011 IEEE symposium on security and privacy 2011 May 22 (pp. 447-462). IEEE.

[42] Vishwanath A, Herath T, Chen R, Wang J, Rao HR. Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. Decision Support Systems. 2011 Jun 1;51(3):576-86.

[43] Bilge L, Sen S, Balzarotti D, Kirda E, Kruegel C. Exposure: A passive dns analysis service to detect and report malicious domains. ACM Transactions on Information and System Security (TISSEC). 2014 Apr 1;16(4):1-28.

[44] Abdelhamid N, Ayesh A, Thabtah F. Phishing detection based associative classification data mining. Expert Systems with Applications. 2014 Oct 1;41(13):5948-59.

[45] Mohammad RM, Thabtah F, McCluskey L. Predicting phishing websites based on self-structuring neural network. Neural Computing and Applications. 2014 Aug;25:443-58.

[46] Arachchilage NA, Love S. Security awareness of computer users: A phishing threat avoidance perspective. Computers in Human Behavior. 2014 Sep 1;38:304-12.

[47] Kumar BS, Ravi V. A survey of the applications of text mining in financial domain. Knowledge-Based Systems. 2016 Dec 15;114:128-47.

[48] Chiew KL, Tan CL, Wong K, Yong KS, Tiong WK. A new hybrid ensemble feature selection framework for machine learning-based phishing detection system. Information Sciences. 2019 May 1;484:153-66.

[49] Mahdavifar S, Ghorbani AA. Application of deep learning to cybersecurity: A survey. Neurocomputing. 2019 Jun 28;347:149-76.

[50] Van Eck NJ, Waltman L. Citation-based clustering of publications using CitNetExplorer and VOSviewer. Scientometrics. 2017 May;111:1053-70.

[51] Van Eck NJ, Waltman L. VOSviewer manual. Manual for VOSviewer version. 2011;1(0).

[52] Van Eck NJ, Waltman L. CitNetExplorer: A new software tool for analyzing and visualizing citation networks. Journal of informetrics. 2014 Oct 1;8(4):802-23.

[53] Mas-Tur A, Roig-Tierno N, Sarin S, Haon C, Sego T, Belkhouja M, Porter A, Merigó JM. Co-citation, bibliographic coupling and leading authors, institutions and countries in the 50 years of Technological Forecasting and Social Change. Technological Forecasting and Social Change. 2021 Apr 1;165:120487.

[54] Cancino C, Merigó JM, Coronado F, Dessouky Y, Dessouky M. Forty years of Computers & Industrial Engineering: A bibliometric analysis. Computers & Industrial Engineering. 2017 Nov 1;113:614-29.

[55] Van Eck NJ, Waltman L. Manual for VOSviewer version 1.6. 10. Leiden: CWTS Universiteit Leiden. 2019 Jan.