Journal of Cybersecurity, 2026, **12**(1), 1–25
https://doi.org/10.1093/cybsec/tyag007
Research paper

Four decades of security and privacy research: evolution of topics, impact, and the community

Nurullah Demir [1,2,*], Christian Böttger [1], Henry Hosseini [1,3], Meryem Demir [4], Christian Wressnegger [5], Norbert Pohlmann [1], Tobias Urban [1]

[1]Institute for Internet Security, Westphalian University of Applied Science, Neidenburger Str. 43, Gelsenkirchen, 45897 North Rhine-Westphalia, Germany
[2]Department of Computer Science, Stanford University, 353 Jane Stanford Way, Stanford, CA 94305, United States
[3]Department of Information Systems, University of Münster, Leonardo-Campus 3, Münster, 48149 North Rhine-Westphalia, Germany
[4]Developmental Psychology, Ruhr University Bochum, Universitätsstraße 150, Bochum, 44801 North Rhine-Westphalia, Germany
[5]KASTEL Security Research Labs, Karlsruhe Institute of Technology, Kaiserstraße 12, Karlsruher, 76131 Baden-Wuerttemberg, Germany
*Corresponding author. Institute for Internet Security/Westphalian University of Applied Sciences, 45897 Gelsenkirchen, Germany.
E-mail: demir@internet-sicherheit.de

Abstract

As digital technologies become increasingly embedded in societal infrastructure, IT security and privacy (S&P) have become critical for protecting sensitive information and preserving trust. These domains have evolved from foundational security measures to address complex challenges introduced by artificial intelligence, regulatory frameworks, and decentralized technologies. This paper presents a longitudinal analysis of the evolution of IT S&P research from 1980 to 2023, analyzing over 13k papers from the most relevant venues. Employing the frameworks of established theories from social sciences, i.e. Latour's actor–network theory, and Bourdieu's forms of capital, along with Leydesdorff's key dimensions in scientometrics, we discuss the evolution of research topics and highlight research priorities in the past and today. We apply modern natural language processing techniques to build a taxonomy of research topics within the S&P community. Using this taxonomy, we analyze the community's thematic development, tracing its growth from 5 topics in the 1980s to 100 distinct research topics, reflecting the field's expanding scope and complexity. Analyzing 0.5M authors, we demonstrate strong collaboration networks in the IT S&P community. We also demonstrate that the proportion of female authors in this community has remained relatively constant over the decades, despite an increase in their research activity in recent years. Finally, we assess factors impacting paper citations, author networks, and the linguistic evolution of the community. This study enhances the understanding of the S&P research community, providing valuable insights into future directions. The data underlying this article, including the analysis code and data processing pipeline, are available in the repository at: https://pulse-of-cybersecurity.com/, which also provides an interactive webpage for exploring our results.

Keywords cybersecurity, scientometrics, bibliometrics, science of science, topic modeling, coauthorship networks, metascience, IT security and privacy taxonomy

Introduction

The evolution of IT security and privacy (S&P) research has closely mirrored advancements in computing and digital communication [1]. In the 1960s, as computers became more widespread, research focused on securing them against unauthorized access [2]. Concurrently, the digital landscape began to present new challenges and threats. The growth of specialized IT S&P venues signals the field's advancement, leading to a notable increase in scholarly papers. Initially, related work appeared at general computing conferences [2]. The establishment of specialized platforms, such as the *IEEE Symposium on Security and Privacy* (IEEE S&P) in 1980 and the *Annual Computer Security Applications Conference* (ACSAC) in 1985, signified a pivotal change in the field. The continual growth of specialized venues, such as the *USENIX Security Symposium* and the *Network and Distributed System Security Symposium (NDSS)*, has facilitated this diversification, allowing researchers to delve deeper into the nuances of emerging threats, reflecting the field's dynamic nature.

In this work, we utilize a novel approach by combining the theoretical framing of Latour's actor–network theory (ANT) [3] and Bourdieu's field theory [4] to assess the S&P research community along three dimensions: (1) cultural capital (i.e. research topics and published papers) (2) symbolic capital (i.e. citation counts), and (3) social capital (i.e. collaboration in the community). This scientometric and metascientific (science-of-science) study contributes to understanding the S&P research landscape and provides insights into collaboration patterns. The objective of this pa-

Received: 22 May 2025. **Revised:** 19 January 2026. **Accepted:** 18 February 2026

per is to examine the evolution of the community from its origins as a niche topic to its current status as a specialized and rapidly expanding field. For this purpose, we examine the growth and evolution of IT S&P research since 1980. Our empirical observation is based on a large-scale analysis of 13k papers from the 11 most relevant S&P venues. We assign a topic to each paper using BERTopic [5] and use additional meta-information (e.g. authors, venues, and citation counts) to perform an in-depth analysis of the S&P research community. With our study, we aim to provide an in-depth overview of the evolution and current state of the S&P research community. Thus, we do not aim to identify governance failures or epistemic blind spots, nor to provide normative directions for the S&P community. On the contrary, this work is a starting point for future science-of-science work in the S&P arena to analyze, for example, why gender imbalances persist in the community or how funding bodies can impact innovation in this field.

The main contributions of this paper are:

- Taxonomy of S&P research topics: We build a taxonomy of research topics in the S&P community (i.e. cultural capital). Based on the taxonomy, we present an overview of the evolution of these topics within the community from a mere handful to over 100 over the past four decades, reflecting the expanding scope and complexity of S&P challenges.
- Revelation of citation factors: We provide an overview of the factors that impact citation counts (i.e. symbolic capital). Our analysis indicates that key factors, including publication in A* venues, the number of coauthors, and engagement with popular or emerging topics, significantly impact citation counts. In contrast, broader topic diversification negatively affects these counts.
- Insight into the collaboration in the community: We provide an overview of the evolution of gender diversity and collaboration networks in the community (i.e. social capital). The findings reveal that the S&P community has been, and remains, male-dominated, with little change over time. Furthermore, we shed light on career longevity patterns and topic migration trends.

Theoretical background

A research community comprises a network of authors, venues, countries, universities, research centers, and institutions [6]. Theoretical frameworks established in the field of scientometrics (i.e. "the science of measuring science") identify three key dimensions for the analysis of a dynamic research community: actors (i.e. authors), venues (e.g. conferences), and outputs (i.e. publications) [7]. Given these key dimensions, we employ two established theoretical frameworks from the sociological sciences [3,4] designed to facilitate understanding communities and their social structures. These frameworks allow us to interpret our findings within a broader understanding of the S&P research field.

Our primary framework is Latour's ANT [3], which, compared to other sociological theoretical frameworks, allows the interpretation of interactions between living and nonliving entities (actors) as an extensive, symmetric, heterogeneous network. This approach allows us to observe the dynamic interactions and dependencies between the named key dimensions of research communities rather than merely examining human interactions, as is common in other social theories.

We combine this framework with Bourdieu's field theory [4]. This theory offers a theoretical lens for interpreting the power dynamics between individuals and their ingrained habits and preferences, shaped by socialization (habitus), within a structured network of relationships among individuals who share common interests and practices (field). The concept of capital in this theory, which encompasses four categories: social, economic, cultural, and symbolic, is of particular interest to our research. According to Bourdieu, humans accumulate these forms of capital over their lifetimes, leading to greater social prestige and recognition.

Social capital refers to having a strong collaboration network, cultural capital to knowledge of popular topics in the S&P community, and symbolic capital to publication in higher-ranked venues and citations. These capital forms can be converted to one another, which fits the context of our research. For example, a stronger collaboration network can lead to a publication in a higher-ranked venue and vice versa. In our analysis, we exclude economic capital. While universities' total budgets are often publicly available, including the shares of state and third-party funding, departmental budgets are rarely disclosed. Moreover, for publicly announced research grants, institutions often collaborate in grant writing, making it infeasible to determine the share of the funding obtained, since the respective proposals are not publicly available. It is crucial to note that research institutions with substantial economic capital benefit from access to tools and resources for superior research, including the acquisition of costly materials and more research staff.

Related work

Our study extends previous work that examined the evolution and trends in S&P research. We use the term "*security*" synonymous for "IT security."

S&P research

Research in S&P has grown in recent years. Suryotrisongko *et al*. [8] reviewed security research topics, taxonomy, and challenges during 2014 and 2019 and analyzed 99 papers. Baset and Denning [9] analyzed the progression of 95 topics and their relationship to authorship distribution in publications from the four A* S&P venues through 2015. Dhawan *et al*. [10] analyzed global security literature and discovered underlying trends and developments from 1998 to 2019. They demonstrated that cybersecurity research increased by 46% in that period. Katsikeas *et al*. [11] identified over 90k active authors in the cybersecurity research area since 1949 and identified different subcommunities in the field. Further, Reuter *et al*. [12] analyzed the evolution of usable S&P research, emphasizing the importance of transparency and tailorability in user interventions. Most related to our work is the System Security Circus, which provides useful high-level figures on the evolution of the S&P community [1]. The website analyzed the top security venues over the past 20 years, providing information on top authors, institutions, and the number of published papers over time. For this reason, we exclude analyzing the individual contributions of institutions to the S&P research community.

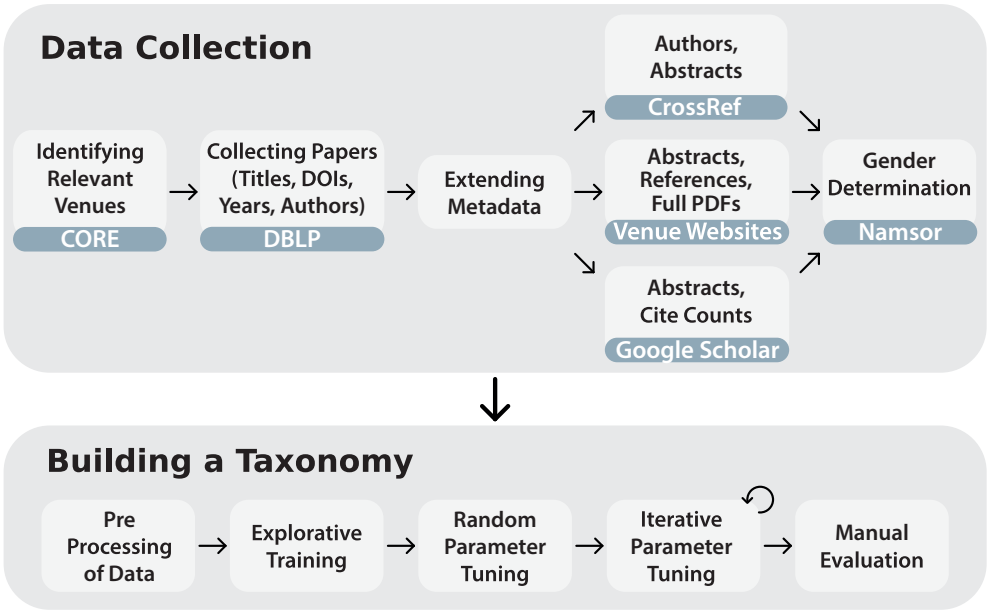

Figure 1 Our collection, data processing, and model training pipeline.

Scientometrics and cybersecurity research

In recent years, scientometric analysis has examined cybersecurity research, focusing on its growth and collaboration patterns. Loan *et al*. [13] analyzed cybersecurity publications from 2011 to 2020, noting increased collaboration among 93 countries and identifying the USA and China as top contributors. Lee *et al*. [14] applied coword analysis to track shifts in IT Security research, demonstrating how themes adapt to new technology needs. Alqurashi *et al*. [15] focused on malware detection, mapping six key research clusters and highlighting the role of interdisciplinary links in cybersecurity. Dhawan *et al*. [10] examined cybersecurity output from 1998 to 2019, highlighting the influence of international collaborations, top countries, and funding on research impact. Wendzel *et al*. [16] analyzed the impact of different factors (e.g. length of paper, publication venue, or authors) on the citation of a paper in different subareas of S&P. The authors find, for example, that the number of references in a paper influences its citation count and that differences exist across subareas. Olijnyk [17] analyzed the differences in papers published in information security based on the authors' affiliation country (e.g. USA or China), the top researchers in the community, and topics (based on keywords) discussed in the community. The author found that China has surpassed the USA in academic output in the field and that the S&P community has been predominantly technical. Lang [18] conducted a bibliometric analysis of 6065 S&P publications using Web of Science data, mapping 20 academic disciplines and identifying key thematic clusters, authors, and journals. The study highlighted the field's fragmentation and underscored the need for more cohesive, multidisciplinary approaches to address the complex, evolving nature of cybersecurity threats and emerging topics effectively. Koca and Çiftçi [19] conducted a bibliometric analysis of 3354 publications from the Web of Science Core Collection between 1975 and 2025 at the intersection of big data and cybersecurity, mapping intellectual structures, thematic clusters, and international collaborations. Their findings highlight the dominance of conference proceedings, the central roles of China and the USA, key thematic trends such as IoT, AI-based intrusion detection, and privacy-preserving technologies, and expose both strengths and structural gaps that limit interdisciplinary research.

Distinction from previous work

This study differs from previous work in several ways. First, it covers the entire span of S&P research, providing insights from its inception to the present. Second, while related studies often rely on keyword analysis to identify trends, our study employs topic analysis, allowing us to identify core topics and organize them into a structured taxonomy of research areas. Third, we integrate citation and community analysis with topic modeling, giving a clearer view of influential works and collaborative networks. Finally, our study also examines shifts in community structure and language within the S&P field, adding context to the evolution of research topics and framing over time.

Data collection

We collected openly accessible data for our analysis between 23 December 2023 and 14 February 2024. The methods and artifacts we outline in this section are publicly available (see the section "Code and data availability"). Figure 1 provides an overview of the data collection process and the development of the taxonomy for topics associated with S&P research, which we detail in the following sections.

Identifying relevant venues

The initial step of our method is the identification of established and well-renowned venues (e.g. conferences) that publish S&P-related papers. To identify these venues, we use the `CORE2023` [20] ranking and focus only on venues in the Cyberse-

Table 1 Analyzed S&P venues, sorted by first appearance. Note: The table lists the venues' current names.

ID	Venue	Abbrreviation	Rank	First	Number of papers
1	IEEE Security and Privacy	IEEE S&P	A*	1980	1759
2	Annual Computer Security Applications Conference	ACSAC	A	1985	1529
3	IEEE Computer Security Foundations Symposium	CSF	A	1988	817
4	European Symposium on Research in Computer Security	ESORICS	A	1990	1315
5	USENIX Security	USENIX Sec	A*	1990	1695
6	Network and Distributed System Security	NDSS	A*	1993	889
7	ACM Conference on Computer and Communications Security	CCS	A*	1993	2405
8	International Symposium on Research in Attacks, Intrusions and Defenses	RAID	A	1998	552
9	Privacy Enhancing Technologies Symposium	PETS	A	2001	846
10	ACM ASIA Conference on Computer and Communications Security	AsiaCCS	A	2006	1085
11	IEEE European Symposium on Security and Privacy	EuroS&P	A	2016	335

curity and Privacy (4604) category, ranked as A* or A. We exclude cryptographic venues, as their focus on theoretical and mathematical security differs from the broader practical and applied S&P topics. Furthermore, we exclude journals, as the CORE journal ranking list has been discontinued since 2022, and the Scimago ranking (SJR) does not contain a dedicated category for cybersecurity and privacy journals. We acknowledge that this limitation may reduce the generalizability of our findings, particularly in research areas primarily driven by journals. Notably, the cybersecurity and privacy community is conference-driven in terms of publications, with the so-called "top 4" conferences (USENIX, CCS, S&P, and NDSS) attracting the most attention [9,19].

We exclude venues that recently introduced a specific focus on S&P. Table 1 lists the analyzed venues.

Identifying and collecting papers

Analyzing the most relevant S&P papers is crucial to understanding the community's activities and contributions. Their metadata offers insights into the research landscape, revealing the scope and focus of scholarly efforts. In our study, we examine metadata elements such as titles, abstracts, author names, or affiliations. The following details our data collection and processing approach.

Digital Bibliography & Library Project

We use the *Digital Bibliography & Library Project* (DBLP) [21], an open-access platform for computer science bibliographies, to extract papers published in relevant venues. After manually identifying the venue pages in DBLP, we developed a crawler to automate the extraction. We collected all listed papers from the earliest available edition through 2023. Since not all venues have published their proceedings when writing this paper, we excluded 2024. Although our dataset is comprehensive, it omits the first proceeding(s) of ACSAC (4), USENIX Sec (3), NDSS (2), RAID (1), and PETS (1), as they are not listed on DBLP. The data collected from DBLP includes (1) paper titles, (2) digital object identifiers (DOIs), (3) tracks of the papers published, (4) locations of the venues, and (5) years of publication. We collect additional information provided by DBLP on the identified papers and author, using the provided API.[1] Some author names are disambigious (e.g. Yang Liu). For these names, the DBLP has specific pages that highlight the disambiguation of authors with the name (e.g. https://dblp.org/pid/51/3710.html for persons with the name Yang Liu). Therefore, some data points in our database do not describe a specific author; instead, they show the complete bibliographies of multiple authors who share the same first and last name. We acknowledge this as a limitation in our work, as this affects 9% of the authors within our scope. We excluded these "authors" from relevant analysis steps (e.g. collaboration networks). To allow a more in-depth analysis, the data was enriched with further information, such as abstracts and affiliations of the authors. To meet this challenge, we used the following additional data sources.

Crossref

We use Crossref [22], an open bibliographic platform, to enrich publication metadata. Through the "Crossref Unified Resource API,"[2] we can query additional metadata using the DOI of each paper (e.g. abstracts, affiliations, and cited references). However, Crossref has limitations that impact our study: (1) the completeness of retrieved data is not guaranteed (e.g. missing abstract), and (2) if a paper does not have a DOI, the API cannot find it. In our study, this primarily affects papers from USENIX Sec, so we use Google Scholar instead.

Publisher's websites

We crawl the webpages of the publishers of the analyzed venues using the DOI URLs provided by DBLP, which redirect to the publishers' sites (i.e. Springer [23], Institute of Electrical and Electronics Engineers (IEEE) [24], or Association for Computing Machinery (ACM) [25]). From these sites, we extract the full PDF of each paper and metadata. Since 2022, the Proceedings of the Privacy Enhancing Technologies Symposium (PoPETS) have been self-published [26]. Thus, we crawled the required metadata for this venue from the venue's site.

[1] https://dblp.org/pid/{author-id}.xml
[2] https://api.crossref.org/works/doi-of-a-paper

Google Scholar

The previous data sources provide the information most papers require. However, we need to use different sources for papers that do not have a DOI (e.g. USENIX Sec papers) as they are not listed or cannot be found using the previously described APIs and services. To address this challenge, we use Google Scholar [27]. As of this writing, Google Scholar does not provide an API for querying publications. Thus, we developed a crawler to extract the abstracts, the number of citations, and the full paper (i.e. PDF file) of all papers. To ensure we only find the desired papers, we use Google Scholar's exact phrase search and then visit the first result returned. Our manual verifications showed that Google Scholar consistently returned correct results. We then store the found paper and all provided metadata.

Identifying the gender of authors

Similar to other STEM fields [28], the computer science community is male-dominated [29]. We aim to analyze if and how this imbalance affects the community along the different dimensions (i.e. social, cultural, and symbolic capital). Thus, we need to assess gender diversity in our community and its evolution over recent decades by first determining the authors' genders. Therefore, we use the authors' full names as stated in the analyzed papers as input to the name-to-gender determination service Namsor [30]. Prior research has indicated that Namsor is effective for accurately mapping names to genders [31–34]. Before classification, we removed any numbers or other residuals following the name (e.g. `001`, which may occur when multiple authors share the same name). We use Namsor's "Genderize Full Name" feature, which takes a name as input and returns a probability for the gender associated with the name.

We applied a 75% confidence threshold for this probability, focusing only on names where Namsor shows relatively high certainty. This cautious approach accounts for the fact that names can refer to multiple sexes and vary across cultures and regions. The commonly associated sex of a name might not always reflect a person's lived identity. Furthermore, analyzing sex based on names raises important ethical and social concerns. First, it inherently reinforces a binary framework (male/female) that excludes nonbinary and gender-diverse identities, as many individuals do not identify with the sex assigned at birth, and their names may not reflect their gender identity. In particular, transgender and nonbinary individuals may face misclassification, which can contribute to identity erasure in research.

Despite these concerns, we used this approach to highlight biases in the S&P community regarding the under-representation of some groups and identities. Furthermore, we want to emphasize that alternative means of determining an author's lived gender identity at the time a paper was published are not feasible. Thus, to better reflect and observe the diversity of a community, authors may disclose their lived gender identity when a paper is published, as indicated in the manuscript metadata. Finally, our goal is not to assess the reasons why gender inequality persists in the community, but rather to make this inequality, to some extent, measurable. Furthermore, our work can be viewed as a first step toward enabling an analytical investigation that helps estimate which measures have reduced inequality. Acknowledging the limitations, our approach is scientifically accepted in other fields [35–37] and has been validated for this purpose [38,39]. As expected, we found that the majority of the authors are male (14 918; 79%) and 4068 (21%) are female.

Building a taxonomy

To analyze evolving research trends, we constructed a standardized taxonomy of S&P topics using topic modeling, an established method for extracting thematic structures from large text corpora. We chose the popular BERTopic modeling technique [5] due to its modern transformer-based approach using Bidirectional Encoder Representations from Transformers (BERT). This method is reported to provide finer topic granularity and enhances the semantic meaningfulness of topics in the context of science-of-science analysis [40,41] compared to traditional topic modeling techniques, such as Latent Dirichlet Allocation [42,43]. BERTopic uses BERT's vector embeddings to generate dense document representations and applies clustering techniques to discover coherent, interpretable topics from large datasets.

We used an unsupervised approach to topic modeling to facilitate a largely unbiased derivation of topics and to enable the construction of topics and the taxonomy at a fine-grained level. This means that some high-level topics (e.g. "privacy" or "formal methods") may not be present due to the absence of seed words, whereas more specific topics (e.g. "Website Fingerprinting and Traffic Analysis") may be present. Thus, the topics are not expected to align with the common conference track names or the keywords provided in the papers.

Data preprocessing

Preprocessing and standardizing collected textual data before training a topic model is essential to reduce noise and ensure consistent input data. This standardization includes expanding common contractions (e.g. "you're" to "you are"), replacing ligature characters (e.g. "fi" to fi"), fixing broken hyphenation caused by line breaks, and lowercasing all texts. We also remove stopwords [44], URLs, and email addresses. Whitespaces are normalized by reducing multiple spaces to a single one. Non-ASCII characters are converted to their ASCII equivalents (e.g. é to e). XML tags and other noncontent formatting elements are removed.

Topic model training and evaluation

Training a BERTopic model is an iterative process, and the lack of standard metrics (e.g. precision or recall) makes it challenging to evaluate the results. We established the following evaluation criteria to identify the best-trained model:

(1) Coherence: The topic coherence measure [45–47] evaluates the degree of semantic similarity between high-scoring words in a topic. This metric indicates the interpretability and relevance of the generated topics, is highly correlated with human judgments [47], and is one of the most important criteria for evaluating the semantic interpretability of topic models [48,49].

(2) Number of topics: It is essential for the model to generate an appropriate number of topics. During our experiments, we noticed the generation of models with relatively few topics (around 10) and models with 70–611 topics. We exclude models with around ten topics, as they are too coarse-grained for our analysis.
(3) Number of outliers: BERTopic sometimes does not assign any topic to documents, leading to outliers. More outliers indicate that the model cannot assign topics to papers, leading to incomplete dataset coverage. To maximize coverage, we opt for models with fewer outliers.

To obtain high-quality embeddings for effective topic modeling, we used the embedding model with the highest rank in semantic search, i.e. `all-mpnet-base-v2` [50,51]. As initial experiments indicated that topic modeling is less effective on the full texts of the papers, we used their titles and abstracts, as these capture the core themes of each paper [52]. Thus, a document is a combination of both elements. We use the *Dynamic Topic Modeling* technique of BERTopic to trace the evolution of topics over time. Our model-building approach consists of the following steps.

(1) Explorative training: In an exploratory study, we generated 30 models using specific parameters for our implementation (e.g. *min_topic_size*, *nr_topics*) and evaluated their performance based on our criteria. The coherence range was 0.3–0.6, indicating low performance. However, manually identifying the optimal parameter space was a time-consuming and challenging process due to the complex nature of the data and the multitude of potential parameter combinations.
(2) Random parameter tuning: Based on the explorative training, we established a systematic parameter space to identify the best parameter ranges. We then trained 5k models with randomly selected parameter combinations to ensure comprehensive coverage and avoid bias, balancing depth of exploration with computational feasibility. In this stage, we achieved a maximum coherence value of 0.7. We observed and determined the model parameters based on our evaluation criteria, and adjusted the parameter space to improve coherence.
(3) Iterative parameter tuning: We developed an efficient approach to calibrate parameters. We proceed with the best model identified in the previous stage and systematically adjust a single parameter at a time, testing whether the coherence value increases, the number of topics remains within a reasonable range, and the number of outliers decreases. If a parameter improves without compromising other metrics, we adopt it and continue iterating on the remaining parameters. Following this method, we train over 7k models, refining the parameter space to improve performance.

Manual evaluation of topic models

We recorded the previously named parameters for all trained models and screened them as follows: (1) select the models with the highest coherence value, (2) exclude models with <10 topics (see the section “Topic model training and evaluation”), (3) sort the models according to the number of outliers, favoring those with the fewest outliers.

The two best models were as follows: the first had a coherence score of 0.76, 100 topics, and 188 outliers. The second model had a coherence score of 0.72, with 127 topics and 194 outliers. Coherence alone does not fully capture the accuracy of topic representation within the original documents [43]. To address this, we manually analyzed 30 randomly selected papers to evaluate the performance of the models, as suggested in related work [49]. Three authors annotated whether a determined topic matched the given abstract. The annotators did not check whether another topic was a better fit, but they did verify that a given topic fit the abstract.

After evaluating the models, we discussed instances in which we reached divergent conclusions (six cases for the first model and seven for the second). Five and four inconsistent options could be resolved. Subsequently, we computed Fleiss' Kappa (κ) to measure the inter-rater agreement. Fleiss' Kappa indicated an “almost perfect” interpretation for both models ($\kappa = 0.91$ for the first model and $\kappa = 0.83$ for the second model). On average, the first model correctly classified 84% of the papers and the second model 81%. Given the second model's lower inter-rater agreement and lower classification accuracy, we perform an in-depth analysis of model one. Therefore, each researcher manually evaluated 100 additional papers to verify the model's performance, resulting in 330 annotated papers. We found an average accuracy of 85% for the assigned topics. For reproducibility [53], we note that the BERTopic model was trained using key parameter settings (`min_topic_size`: 10, `min_cluster_size`: 30, `n_neighbors`: 10, `ngram_range`: (1, 3)); the complete set of parameter values is provided in the Supplementary materials (see the section “Code and data availability”).

Adjusting topic labels

The initial topic labels were generated by BERTopic, resulting in labels, such as `android_apps_android apps_android applications`, which lack clarity. To enhance their readability, we leveraged ChatGPT to refine these labels [54]. This process transformed them into comprehensible labels (e.g. “Android App Security & Privacy”), rendering the topics more meaningful to a human audience. Afterward, two expert authors optimized the labels by cross-checking them against the documents, ensuring clarity and simplicity in their representation (e.g. modifying “Graph-Based Alert Causality Analysis” to “Causality in Log Data Analysis”). Table 2 lists the top 20 topics, and Table B1 lists all 100 topics identified in the papers, sorted by popularity, illustrating our final taxonomy.

Lexical analysis

We use two established linguistic techniques to examine longitudinal lexical shifts in the language of the collected papers. Keyness analysis identifies significant changes in the occurrence of phrases over time. The significance is determined using the 4-term log-likelihood estimation (G^2) [55] and the Bayesian Information Criterion (BIC) [56], calculated as $BIC = G^2 - \ln(N)$, where N refers to the sum of tokens observed in two compared subcorpora. The subcorpora comprise the same data used for training the topic model and are grouped by the longitudinal topical trends described in the section “Topic trends over time.”

Table 2 Identified topics in S&P research ordered by the number of papers in each topic.

ID	Topic name	$\sum$	ID	Topic name	$\sum$
0	Processor side-channel attacks	469	1	Android apps security and privacy	448
4	Cryptographic security protocol analysis	423	3	Blockchain security and privacy	414
9	Control flow integrity and code reuse attacks	361	2	Botnet detection and analysis	337
6	Access control security policies	320	5	Tor and network anonymity	313
8	Intrusion detection systems	278	7	GDPR and data privacy practices	252
13	Trusted computing and other trusted systems	249	10	Adversarial attacks on neural networks	248
21	Secure signature schemes and protocols	236	11	Differential privacy mechanisms and guarantees	232
14	Secure multiparty computation techniques	231	18	Malware detection and deobfuscation techniques	230
15	Web security and browser vulnerabilities	226	12	(Distributed) Denial of service attacks	218
24	Secure information flow control	205	17	Biometric security in digital devices	189

We use collocation analysis to determine changes in the collocates of frequently used terms over time. The window size (left and right) and the cutoff size were set to 5, and the 10 collocates with the strongest association based on the LogDice score [57] were considered for each lustrum and term. The LogDice score offers the advantage of a fixed range (0–14) and independence of corpus size. Similar applications of these methods in the S&P community include detecting change points in the terminology of privacy policies over time [58] and analyzing their longitudinal content [59].

Used statistical methods

In this study, we use various statistical methods to analyze our data. First, the *Kruskal–Wallis* test is used to compare the number of research topics across different venues. Second, the Spearman rank correlation helps us assess monotonic associations between variables such as publication volume and topic diversity. Third, we calculate *Fleiss' Kappa* to measure the reliability of manual topic evaluations. Fourth, we apply a *4-term log-likelihood* (G^2) test combined with the *BIC* in our keyness analysis to detect significant changes in vocabulary over time. Finally, *Structural Equation Modeling (SEM)* with robust maximum-likelihood estimation is employed to explore interdependencies between factors, such as venue rank, coauthorship, and citation metrics. All tests are carried out with a significance level of $\alpha = 0.05$.

Research output trends (cultural capital)

In our study, we analyze the three key dimensions for research community analysis (authors, venues, and outputs) [7]. We present research trends (i.e. cultural capital) observed within the S&P community from 1980 to 2023 to assess the output of the S&P research community.

Trends across venues

First, we present the high-level results derived from our dataset. We collected 14 968 publications, including posters, editorial notes, panel discussions, keynote talks, and workshop papers. We excluded all entries that were not full papers and papers for which we could not collect an abstract. After filtering, our analysis dataset comprised 13 134 (88%) full papers. Figure 2 shows the number of papers published annually across the analyzed venues. Appendix A shows the distributions of published papers by year. An exponential growth model indicates an annual increase of 11% (min: −41%, max: 165%, SD: 32%). The dataset shows a substantial increase in publications, from an initial count of 19 to 1548 over the study period, highlighting significant growth in academic output. However, this increase may also be influenced by the inclusion of new venues, which can inflate paper counts. Thus, we normalize the data for further analysis by accounting for the number of venues each year.

Due to the exponential growth of publications, we split the papers into three periods based on the growth rates to analyze the evolution of the community at different times. From 1980 to 2003 (early years), the normalized annual growth rate in published papers was low, at about 1%, with an extensive range of year-to-year changes (min: −26%, max: 77%, SD: 21%), reflecting a gradual increase in research activity during this period. Between 2004 and 2013 (middle years), we observe a rise in this growth rate to 7% (min: −2%, max: 18%, SD: 6%), suggesting a phase of stronger growth and more activity within the research community. The recent period from 2014 to 2023 (current years) shows a higher growth, with a normalized annual rate of nearly 12% (min: −3%, max: 28%, SD: 9%), reflecting a consistent increase in the number of publications. CCS, IEEE S&P, USENIX Security, and ESORICS have seen the highest increases over the past 4 years. These figures reflect a robust and increasing engagement in S&P research. Overall, CCS (2544), IEEE S&P (1264), and USENIX Security (1823) have the highest number of published papers. In contrast, EuroS&P (231) and RAID (450) have the fewest.

Topic trends over time

As described in Section "Manual Evaluation of Topic Models" and depicted in Table 2, the trained BERTopic model identified 100 distinct topics in our dataset. Figure 3 provides an overview of the topical diversity over the past 40 years. We found a significant relationship between the number of topics and the number of papers ($P < .0001$, Spearman rank $= 0.98$). This finding confirms that the number of distinct research topics increases proportionally with the number of papers. Remarkably, since 2021, the number of pa-

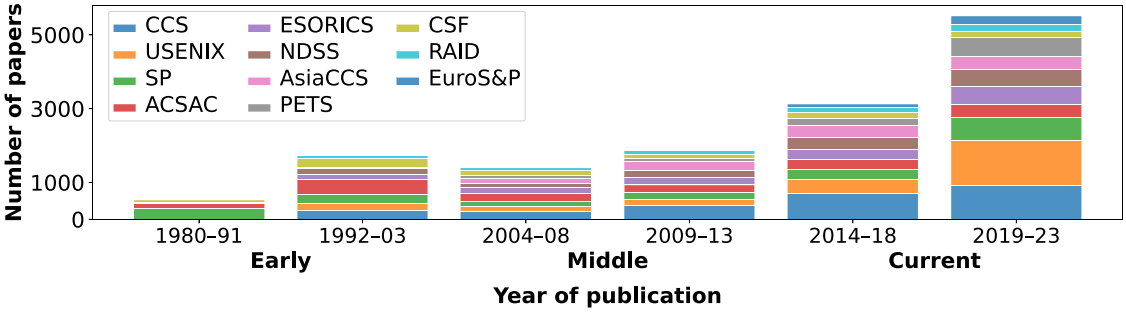

Figure 2 Number of papers by year and venue.

pers has substantially increased without a corresponding increase in topics. The annual growth rate for the number of topics indicates an average increase of 7% (min: −29%, max: 82%, SD: 20%), reflecting considerable year-over-year variability in topic growth.

Table 3 provides the evolving landscape of research topics. A linear regression analysis demonstrates the rapidity (i.e. slope) with which a research area is gaining or losing popularity, as reflected in the number of papers published on each topic. Overall, the trends observed indicate that the community's research interests fluctuate.

Early years (1980–2003)

In the early years of S&P research, emphasis was placed on establishing foundational frameworks. *Trusted Computing and other Trusted Systems* (slope 1.8) and *Information Flow Security Analysis* (slope 1.3) experienced growth between 1980 and 1991, reflecting the need for security systems. Between 1992 and 2003, the rise in *Intrusion Detection System Techniques* (slope 2.4) highlights the growing concern for proactive security measures. These trends underline the initial focus on developing technologies to safeguard emerging digital infrastructures. In this period, we identified 79 topics. In the early years, downward trends were limited, reflecting a period when S&P research was developing its foundational concepts. The slight decline in topics such as *Cryptographic Security and Analysis* (slope −0.1) may indicate a shift from theoretical exploration to practical implementation as cryptographic methods became standardized, alongside the shift of cryptography-related publications to more specialized venues, which are excluded from this study.

Middle years (2004–2013)

This period demonstrates a transition toward combating emerging threats. During 2004 and 2008, *Tor & Network Anonymity* (slope 2.9) and *Malware Detection and Deobfuscation Techniques* (slope 2.1) surged in response to the increasing need for privacy in digital communications and the growing complexity of attack vectors. Between 2009 and 2013, the focus on *Android App Security & Privacy* (slope 7.8) highlighted the shift in security practices to mobile platforms. In this period, we identified 99 topics, a 25% increase from the early years. As S&P research evolved during the middle years, the focus on certain areas such as *Trusted Computing and other Trusted Systems* (slope −1.2) and *Secure Database Management Systems* (slope −0.7) declined. This reduction likely reflects a transition toward addressing emerging challenges that require innovative security solutions beyond traditional methods.

Current (2014–2023)

In the last decade, we have seen a focus on advanced technological challenges. Topics like *Processor Side-Channel Attacks* (slope 8.4) and *Adversarial Attacks on Neural Networks* (slope 10.7) demonstrate the transition toward addressing sophisticated attacks. The substantial increase in publications on *Blockchain Security and Privacy* (slope 4.6) and *GDPR and Data Privacy Practices* (slope 10.2) reflects the response to new regulatory and technological landscapes. In this period, we observed all 100 topics, indicating a minimal increase in the diversity and scale of topics that the S&P community has researched in the near past. In the same period, significant downward trends have been observed in areas like *TLS Certificate Security and Validation* (slope −1.4) and *Trusted Computing and other Trusted Systems* (slope −2.2). These declines likely reflect the maturation of these topics, where foundational research transitions into engineering standardization (e.g. *IETF TLS 1.3* [60]) and industry practice, shifting the community's focus from novel discovery to refinement and deployment [61].

Leading topic trends

This section analyzes the most influential topics from 1980 to 2023. We first present the life cycle of the "leading topics," i.e. the topics with the highest number of papers in at least 1 year, highlighting their emergence and evolution over time. We then focus on the most prominent topics of the last decade and assess their prevalence.

Life cycle of leading topics

We define a topic as *leading* if it was the most popular in any given year (i.e. the topic with the most papers in a year). Overall, we identified 11 leading topics. We examine how these leading topics emerged and evolved in terms of publication volume. Figure 4 provides an overview of the life cycle of these leading topics, revealing their changing prominence over time. At a high level, we observe that the relative dominance of leading topics has declined over time, with no single topic accounting for a large share of papers in recent years. This trend reflects a broadening of the field, with research efforts distributed across a wider set of specialized and emerging areas rather than concentrated in a few.

First, we examine the life cycle of leading topics. In the first 26 years of S&P research, three topics dominated: *41: Secure Database Management Systems* (11 years), *4: Cryptographic Security Protocol Analysis* (5 years), and *8: Intrusion Detection Systems*

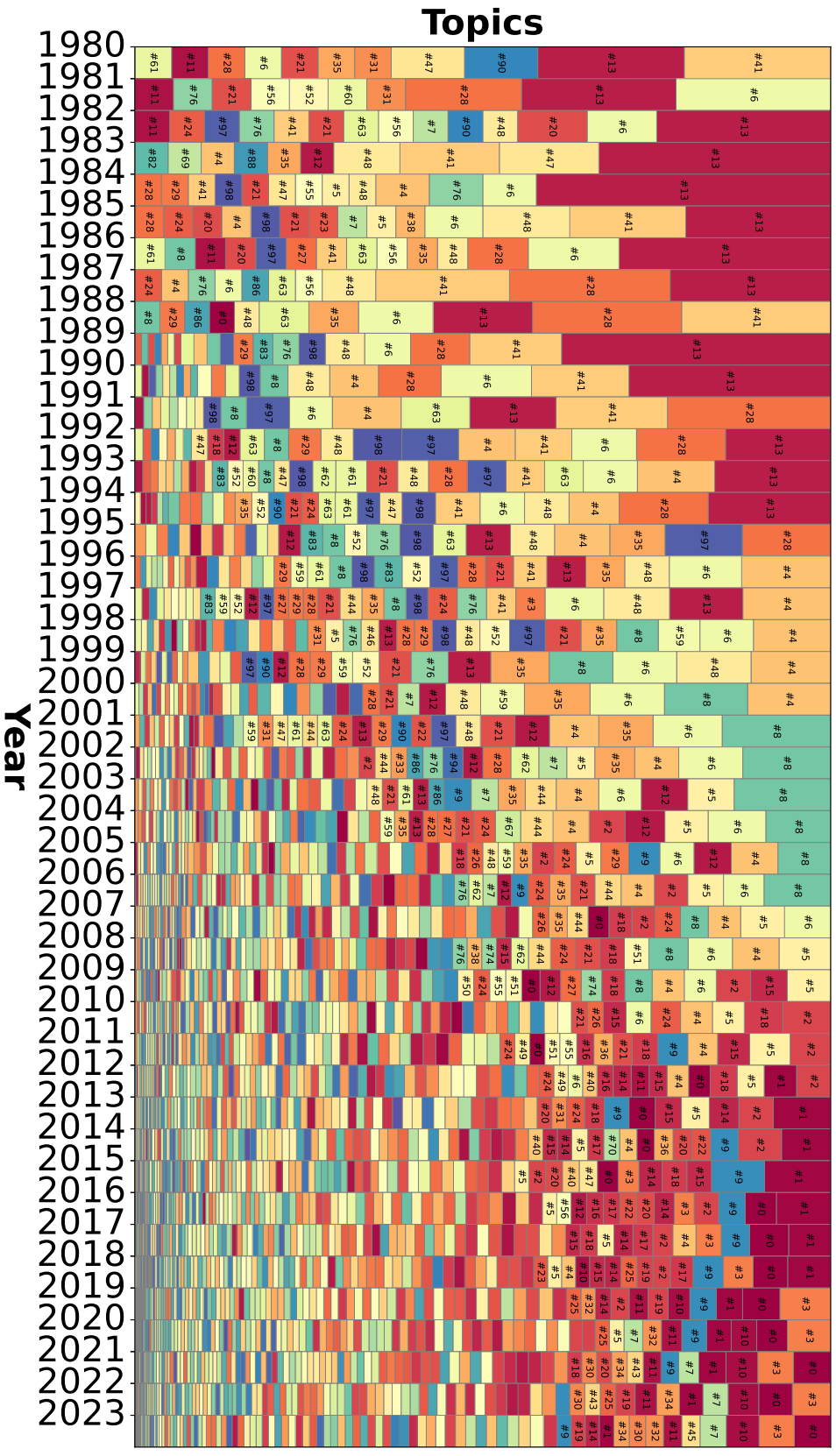

Figure 3 Diversity of the topics over time; longitudinally, the number of topics increased.

Table 3 Top increasing and decreasing topics over the analysis timespan [incline (positive slope), decline (negative slope)].

	Growing		Declining	
	Topic	**Slope**	**Topic**	**Slope**
Early 1980–91	13: Trusted systems	1.8	20: Enc. in DBs	0.0
	28: Information flow	1.3	31: Crypto. vuln.	−0.1
	41: Secure DBMS	1.2	21: Signatures	−0.1
Middle 2004–08	05: Network anonym.	2.9	02: Phishing	−0.9
	18: Malware	2.1	67: Alert causality	−1.3
	51: Rootkit detect.	1.9	12: DoS	−2.0
Current 2014–18	00: Side-channels	6.1	70: WiFi sec.	−1.5
	03: Blockchain	4.6	36: Web apps	−2.0
	17: Biometric sec.	3.3	02: Phishing	−2.2
Early 1992–03	08: IDS	2.4	98: DB concurrency	−0.5
	35: Trust mgmnt.	0.9	41: Secure DBMS	−0.7
	06: Access control	0.8	28: Information flow	−0.7
Middle 2009–13	01: Android sec.	7.8	04: Crypto. sec.	−1.4
	14: Secure MPC	4.1	74: RFID	−1.7
	00: Side-channels	2.3	06: Access control	−2.5
Current 2019–23	10: Attacks neur. net.	10.7	55: Hypervisors	−0.9
	07: GDPR and privacy	10.2	24: Information flow	−1.0
	45: Fed. learn. sec.	8.6	22: TLS security	−1.4

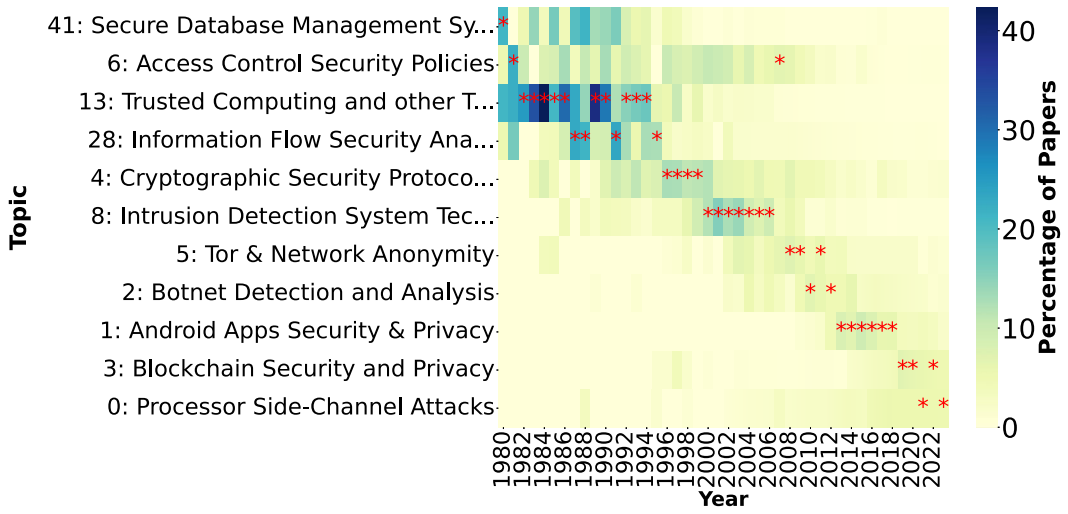

Figure 4 Life cycle of leading topics. * indicates the years when these topics had the most papers.

(6 years). Over the next 18 years, we identified five leading topics, indicating dynamism in this research ecosystem. We find that *Trusted Computing and other Trusted Systems* was discussed for 11 years, from 1983 to 1995 (except for 2 years), spotlighting the importance in early S&P research. This topic was followed by *Intrusion Detection System Techniques* and *Android Apps Security & Privacy*, each leading for 6 years. *Android Apps Security & Privacy* had the shortest trajectory to becoming the leading topic, reflecting its adoption and relevance. The other leading topics have shown a more gradual, balanced increase in papers over the years. Typically, a topic remains leading for consecutive years or with brief gaps. However, *Access Control Security Policies* was a leading topic in 1981 and 2007, with a gap of 26 years, suggesting that evolving technology can reawaken interest in a topic.

Our results demonstrate that the prominence of topics shifts with technological changes, showing the field's adaptability to new challenges. Recurring interest in foundational topics like *Access Control Security Policies* and *Cryptographic Security Protocol Analysis* suggests that fundamental security concepts remain crucial despite technological advancements. The rapid rise of topics like *Android Apps Security & Privacy* highlights the community's responsiveness to emerging technologies. From the perspective of ANT, the growing number of security challenges led to the creation of new security topics, consequently reshaping the diversity

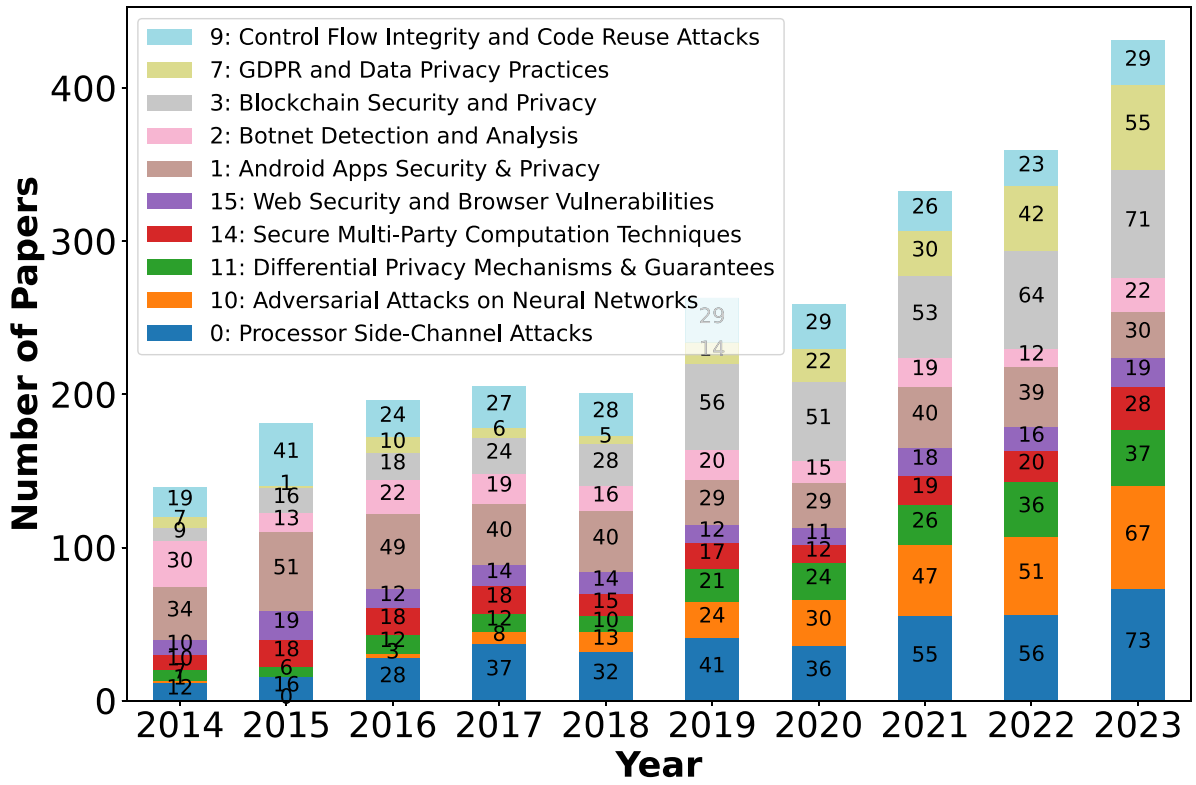

Figure 5 Number of papers on hot topics in the past decade.

of S&P venues. Expanding the range of topics and publications enhances the S&P community's cultural capital.

Leading topics in the last decade

In Fig. 5, we provide an overview of the most prominent topics in the past decade. Our results demonstrate that most papers (417) were on *1: Android Apps Security & Privacy*, followed by *0: Processor Side-Channel Attacks* (407 papers), and *3: Blockchain Security and Privacy* (396 papers). In 2023, the top topics were *Side-Channel Attacks* (78 papers), *Blockchain Security and Privacy* (75 papers), and *Adversarial Attacks on Neural Networks* (70 papers). The prominence of topics like *Blockchain Security and Privacy* reflects the S&P community's efforts to enhance security and trust in decentralized systems. Moreover, in the last 5 years, *GDPR and Data Privacy Practices* has grown the most (400%), followed by *Adversarial Attacks on Neural Networks* (almost 300%) and *Side-Channel Attacks* (245%).

An in-depth analysis reveals that research on *Security of Machine Learning* techniques spans multiple topics (e.g. topics 10, 45, 34, 58, and 80). Combining these topics, we find 68 papers in 2020 and 173 papers in 2023, making it the leading research area recently.

Topic coverage by venues

This section examines the variation of research topics across S&P venues, highlighting their impact on research scope and diversity. Figure 6 demonstrates the number of topics for each venue. The Kruskal–Wallis test ($P < .0001$) [62] indicates statistically significant differences between the venues in the number of topics covered, reflecting varying scopes. Since 2005, the number of topics per venue has consistently increased, highlighting the growing breadth of S&P research in this evolving field.

On average, CCS has 25 topics per year, followed by IEEE S&P (22) and ACSAC (19). As indicated by the *Shannon Index* [63], a metric to measure diversity, CCS, IEEE S&P, and AsiaCCS (4.7 each) exhibit the highest diversity in topics, while CSF (3.0) and PETS (3.4) demonstrate the lowest topical diversity. This observation is reasonable, given that these venues target a specific subfield. Reviewing the average number of topics for the last 5 years reveals that USENIX Security has an average of 63 topics, followed by CCS (62) and IEEE S&P (50). This increase underscores the central role of these venues in S&P research.

First appearance of topics per venue

Here, we aim to gain insight into the emergence of new topics across venues by identifying those that have pioneered them. Our results reveal that IEEE S&P has consistently introduced new topics over the years, with 50 of them, underscoring its pivotal role in shaping the field. CCS (21) and USENIX Security (13) also make notable contributions. AsiaCCS and NDSS introduced only one new topic each, while PETS and CSF introduced two each, reflecting their specialized focus and contributions to subareas. It should be noted that a total of 8 of the top 10 topics were introduced by IEEE S&P.

Linguistic evolution

Changes in vocabulary and the appearance and disappearance of terminology in S&P papers contribute to longitudinal understanding of the field.

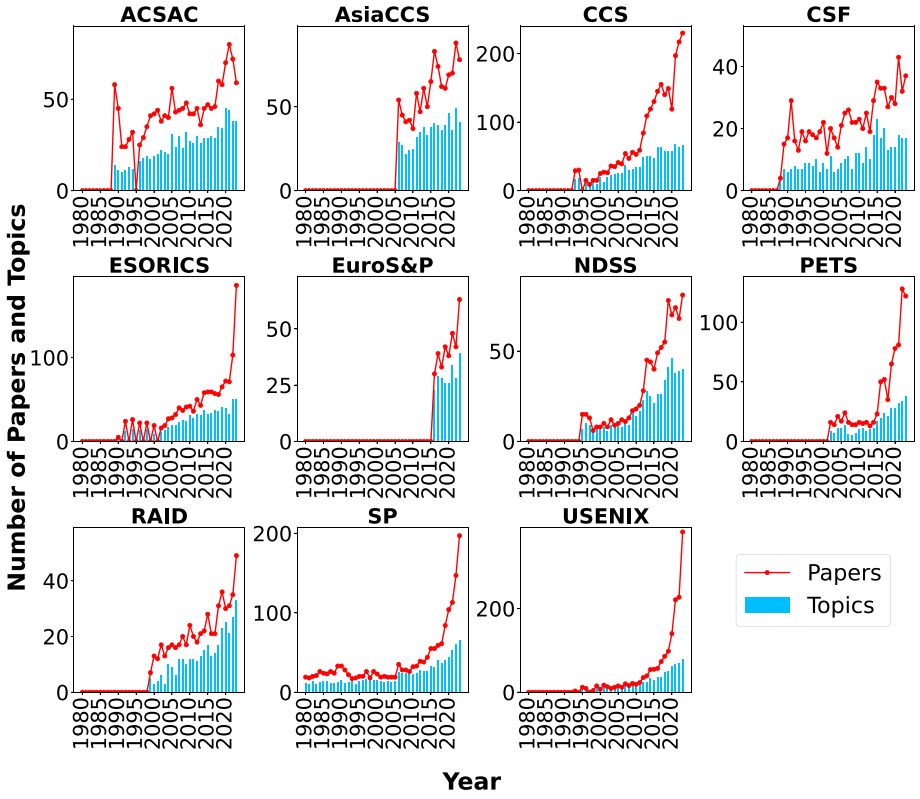

Figure 6 Number of topics over the years across venues.

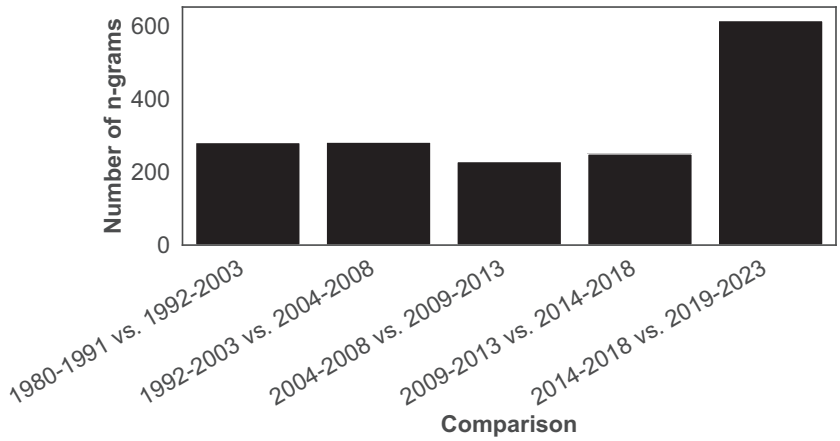

Figure 7 The number of *n*-grams whose occurrence changed with strong evidence between two consecutive topic timespans (see Table 3), demonstrating the ongoing evolution of vocabulary.

Lexical evolution

Shifts in the use of lexemes in scientific publications over time indicate changes in writing style or topic. To identify these shifts, we counted the number of *n*-grams ($n = \{1, 2, 3\}$) whose occurrence increased or decreased significantly (BIC>6) between consecutive topical timespan pairs (see Table 3). Figure 7 demonstrates that fluctuations occurred in the vocabulary and phrases used over time, specifically in the last decade, showing a faster introduction of new terms and the disappearance of others.

Examples of such fluctuations include phrases related to adversarial attacks, GDPR, and federated learning after 2019, similar to the topics identified in Table 3. Terms such as "SSL," "access control," and "tor" have been disappearing from abstracts since 2019, while the occurrence of "fuzzing," "IoT," and "blockchain" has been increasing. We observed stylistic changes regarding paper writing in S&P publications. For example, the occurrence of "in order to," and terms containing "may" and "would" decreased over time, probably related to the growing demand for the conciseness of academic writing.

Collocational evolution

In our second analysis, we selected a set of common terms in abstracts based on their frequency in our documents. Changes in their collocating lexemes symbolize the evolution of language in

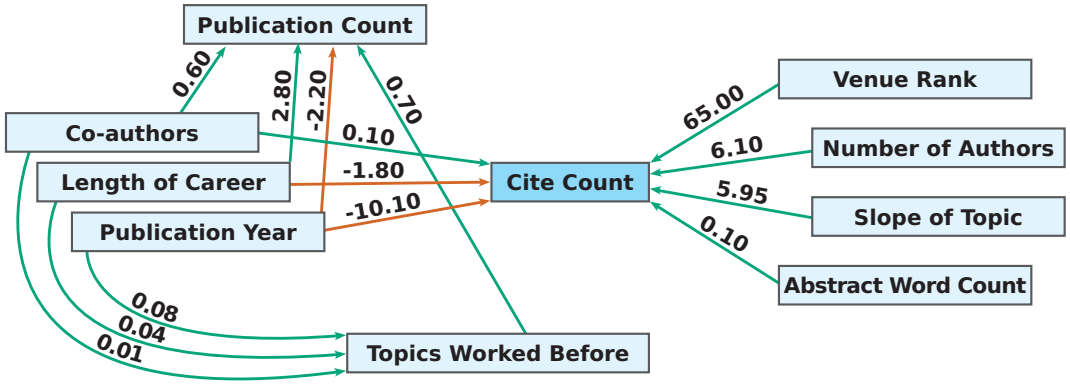

Figure 8 The Structural Equation Model. Nodes represent outcome variables, and edges represent relationships. Green edges (positive weights) indicate positive effects, and orange edges (negative weights) have negative effects.

S&P papers. These terms (nodes) are: *security*, *privacy*, *method*, *propose*, *solution*, *show*, *demonstrate*, and *contribution*.

Our analysis shows a shift toward using "new" and "novel" with the nodes *contribution* and *propose*, underscoring the need to present novelties in top S&P venues to accumulate symbolic capital. Moreover, ordinal numbers have been collocates of *contribution* since 2004, suggesting that contributions are presented in a structured way in S&P papers, thereby indicating a desire to display cultural capital (stylish academic writing) in high-ranked publications. Since the beginning of the 21st century, "provide," "problem," and "propose" have been used almost continuously alongside the node *solution*. This node has also been accompanied by "compare" since 2014, likely due to the need to compare new solutions with previously published ones to demonstrate cultural capital. The node *show* has been collocated with "evaluation" since 2009. Furthermore, "experiment," "result," and "attack," are fixed combinations of this node over the last three decades.

The node *security* has persistently been co-occurring with "system." While "computer" and "kernel" have disappeared as collocates of *security* since 1999, "protocol" has constantly co-occurred with this node since then. Since 2014, *security* has frequently been accompanied by "privacy," suggesting a trend toward including both aspects in recent research topics. "Concern" and "user" have been present alongside *privacy* since 1992, and "preserve" has been constantly present as a collocate since 2004. Although "policy" would have been expected to be continuously present along the node *privacy* due to privacy policy analysis, their combination was present from 2004 to 2008, and between 2019 and 2023, the reason for the latter likely being the recent introduction of impactful privacy regulations such as EU's GDPR or California's CCPA/CPRA, which raised the attractiveness of this topic for privacy researchers. The common collocates of *demonstrate* in recent years, i.e. "effectiveness," "feasibility," "real," and "experiment" likely indicate the increased value of real-world practical solutions.

Research impact (symbolic capital)

Building on our theoretical approach to analyzing the interactions among actors in the S&P research community and understanding the factors that lead to the accumulation of cultural capital, this section examines the factors that influence the academic impact of publications, particularly citation counts and symbolic capital within the community. On average, each paper has 100 citations (min: 0; max: 9132; SD: 287). First, we conduct a statistical analysis using *SEM* to examine how various paper characteristics affect citation rates. Then, we explore the relationship between the development of the topics identified in the previous section and citation counts.

Citation patterns

We use the "lavaan" `R` package [64] to perform SEM [65]. SEM is a method for analyzing relationships among multiple variables to examine their interrelationships. We consider the following features (predictors) for our model: (1) publication year, (2) venue rank, (3) number of authors, (4) topic popularity, (5) topic popularity at publication time, (6) title length, (7) abstract length, (8) number of publications, (9) coauthor network, (10) author career length, and (11) number of topics formerly worked on. Data preprocessing involved handling missing values, identifying outliers (Z-scores ± 3), and fitting the model using robust maximum-likelihood estimation to address non-normality. Our model's R^2 values indicate that the predictors explain 24% of the variance in citation count, 67% in average publication count, and 4% in the number of formerly worked on topics.

Our analysis revealed multiple significant factors influencing citation counts (all P-values $< .01$). Figure 8 demonstrates the predictors significantly affecting the citation count. Only topic popularity and the number of topics formerly worked on did not indicate a significant effect on citation count.

Positive effects

The most influential predictor is the venue rank, with publications in $\mathbb{A}^*$ venues leading to an increase of 65 citations, demonstrating that the cultural capital (research quality) of the authors is converted into the accumulation of symbolic capital and the affordance of *legitimate authority* within the field.

Notably, each additional author on a paper increases the citation count by 6.1, underscoring the significant role of coauthorship in shaping citation metrics and, in turn, the influence of social capital in converting coauthors' cultural capital into recognized publications. As the topical slope, which indicates the trending nature of a topic, increases, citation counts rise by 6, demonstrating that

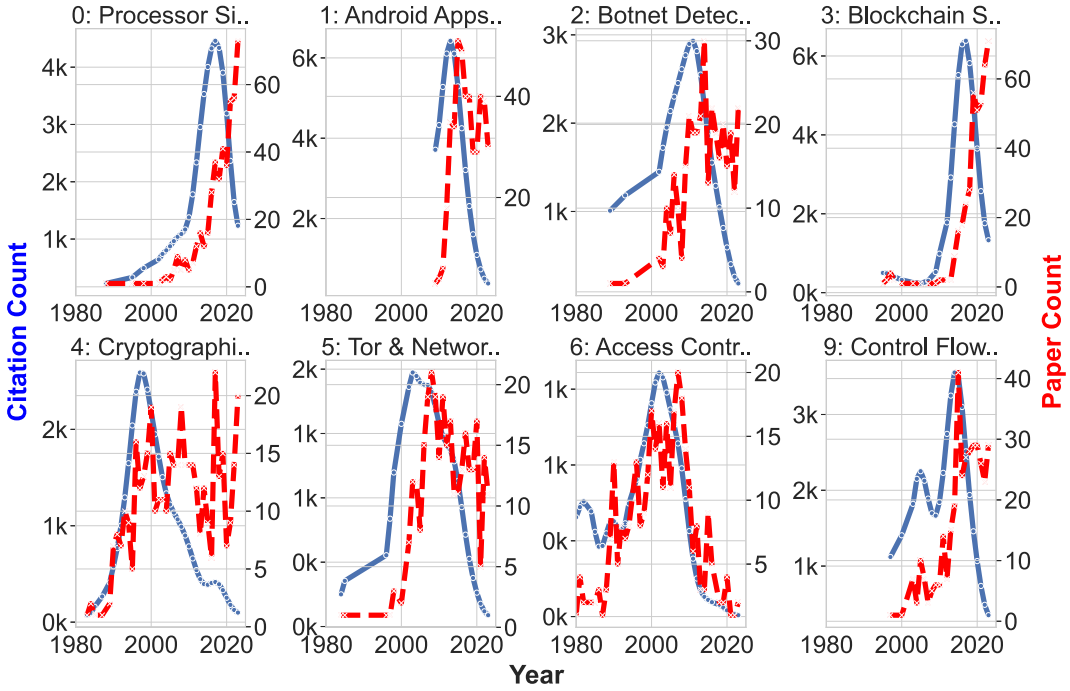

Figure 9 Number of citations and published papers for the top eight topics in the past 44 years.

investing in emerging or growing fields as a form of cultural capital speculation can enhance a paper's impact.

Going from the least trending topic to the most trending topic corresponds to a predicted difference of 17 citations. Additionally, longer abstracts are associated with higher citation counts. This might be due to more detailed abstracts increasing visibility in search engines and attracting more academic interest. Similarly, older papers accumulate more citations over time, gaining 10 citations per year, reflecting the continued relevance and influence of earlier research. Finally, having more coauthors increases the total publication count by 0.6, suggesting that collaborative efforts positively contribute to research output and demonstrating the conversion of social capital into symbolic capital.

Negative effects

Our analysis demonstrates that longer academic careers, the number of words in a paper's title (marginally significant, P-value $<$ .07), and the number of topics an author works on (marginally significant, P-value $<$ .1) negatively affect citation counts. The reason may be that longer titles may not capture readers' attention. Furthermore, in longer careers, researchers might stick to topics that are no longer of broad interest, reducing citation counts. While working on more topics seems to reduce citation counts, it also increases publication volume, positively impacting citations.

Topic popularity and citation counts

Next, we examine the relationship between publication year and citation count. We focus on the correlation between publication volume and citations. Figure 9 demonstrates their relationship for the top eight leading topics. The results show a pattern where papers published early, i.e. when a topic emerges, typically receive fewer citations. Citations peak shortly before the publication volume peaks, suggesting that papers authored during the rising phase of a topic's popularity receive the most attention. Further examination indicates that the peak citation period coincides with the early phase of a topic's development. Thus, the first papers on a topic often receive less attention than follow-up work as the community publishes more on it. However, late publications receive less attention. This highlights the influential role of foundational papers that establish core concepts and methods and drive the topic forward, a pattern observed across almost all analyzed topics. The results show that papers published after the peak are cited less, with citation rates declining after a topic's publication peak.

Diversity in the community (social capital)

This section examines diversity-related factors that contribute to understanding the social capital of authors, who publish within the S&P community. We evaluate the length of academic careers of individuals publishing in the community alongside their associated research topics and output. Subsequently, we focus on the authors' collaboration networks, incorporating their gender to better understand the evolution of the community.

Academic careers and research output

First, we analyze the timespan in which the community generates cultural capital (i.e. research papers) to analyze the authors' social capital within the community. Thus, we analyze the duration of academic research careers of authors who publish in academic S&P venues. We collected all authors (20 691) within our scope to analyze academic career duration with greater precision.

The mean publication period for authors was 13 years (min: 0, max: 67, SD: 10.2, Mdn: 11). The minimum value of zero years refers

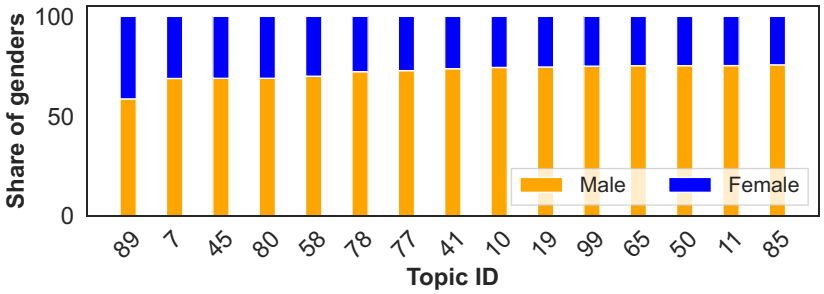

Figure 10 Top 15 topics addressed by female authors compared to the number of male authors active.

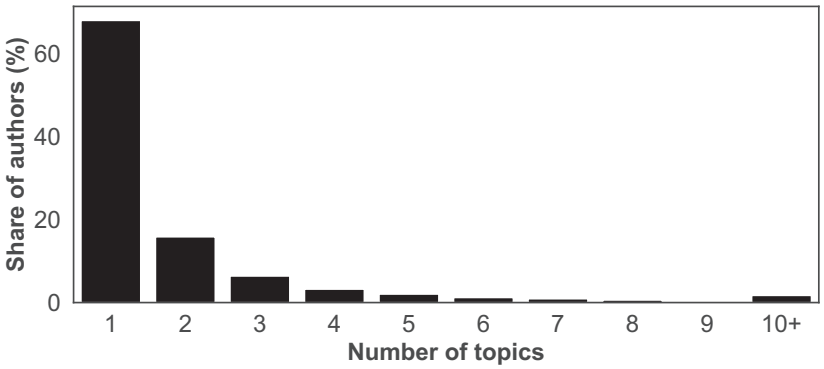

Figure 11 Distribution of the various topics by percentage of authors publishing on them.

to authors who published a single paper. If we limit our consideration to papers published at S&P venues within our scope (see Table 1), the observed "career" durations become shorter. In this case, the average timespan is 2 (min: 0, max: 42, SD: 4.5, Mdn: 0) years, demonstrating that authors active in the S&P community are also active at other communities. The results distinguish between authors, who remain in academia and those who leave or stop publishing (e.g. after a doctorate) by the length of their academic careers [66].

Authors, on average, published at 1 (min: 0, max: 4, SD: 0.7) distinct A* venues during their academic career. 11 603 (61%) authors published in an A* venue. 12 664 (67%) authors published papers exclusively at a single venue throughout their careers. Of these authors, 1245 (6%) published multiple papers at that venue. The authors (7566), who published more than one paper did so, on average, at 3 (min: 1, max: 11, SD: 1.6, Mdn: 2) venues. The lower acceptance rates at A* venues [67] may explain why only approximately two-thirds of authors published there. The results indicate that authors publish across all analyzed venues, with relatively indiscriminate preferences and no particular focus on any venue.

Authors' interest in topics

In the following, we analyze the relationship between authors' social capital and the community's cultural capital, focusing on the relative popularity of topics among authors. Using the identified topic of each paper (one-to-one), as detailed in the section "Building a taxonomy," we examined the topics regarding their author distribution. We distinguish between male and female authors (see the section "Identifying the gender of authors"). Both female and male authors publish on all 100 topics within the scope of this study. However, on average, male authors worked on ˜300% (min: 41%, max: 1033%: SD: 147%) more topics than female authors. Figure 10 provides an overview of the 15 topics with the highest proportion of female (co-)authors. Topic 89 ("User Behavior in Digital Security Awareness") is the most popular with female (co-)authors. Even on these topics, male authorship still surpasses female authorship. This observation again highlights the male dominance in the S&P community.

We identified a significant statistical effect ($P < .0001$) for the gender–topic relationship. This result suggests that female researchers are more active in specific topics, while male researchers cover a broader range of topics throughout their careers. It could also suggest the field's habitus to privilege male-dominated research topics (technical over social topics, e.g. topic 89).

Changing topics in a career

In the following, we present an analysis of the number of topics the authors published about and the frequency with which they changed their topics. 12 893 (68%) of the authors focus on a single topic, while, on average, authors address 2 (min: 1, max: 46, SD: 2.2) different topics. These numbers appear reasonable, given that doctoral students may focus on a single topic and subsequently leave academia, or authors who have published only a single paper may concentrate on that single topic. Figure 11 provides an overview of the number of topics and the share of authors who publish on one or more topics. The analysis shows that most authors publish only on one topic, while only a tiny share publish across multiple topics. The observation is logical, since many au-

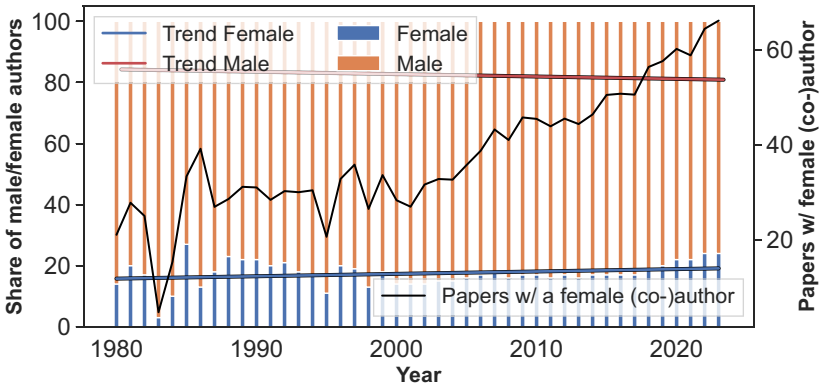

Figure 12 Gender authorship evolution from 1980 to 2023.

thors quit academia after a short period (e.g. after a doctorate) and focus on only one or a few topics during that time. 32% of the authors change or add topics to their research at least once, and 16% do so more than three times. The authors engaged in research on various topics and occasionally shifted their focus, e.g. to explore a new area. From a field-theory perspective, this likely demonstrates the intent to acquire cultural capital.

We focus on the 6092 (32%) authors, who changed their topics at least once to better understand topic changes. We analyze them because they contribute to the community's cultural capital, potentially influencing it through social capital. These authors published 2.7 (min: 1, max: 143, SD: 5) papers per topic. Only 288 (1.5%) authors changed to a leading topic, while 271 (1.4%) newcomers started with such a topic. In an average 6-year period, an author works on 3.7 topics. However, a topic's average lifetime is only 2 years, indicating the pace of the topic changing. An explanation is that researchers with multiple topics work 3–4 years during their doctoral program and 2 years as postdoctoral researchers before leaving academia [68,69].

Collaboration networks

This section presents an analysis of the coauthor networks to demonstrate the extent of collaboration and author cliques within the community. Thus, this section explores the connectivity among actors within the ANT framework and its implications for social and cultural capital. Therefore, we examine all authors within our scope (18 986) and their respective coauthors, resulting in a total of 509 382 authors. On average, an author collaborates with 51 (min: 1, max: 2108, SD: 84, Mdn: 22) other authors during their career. On average, a paper published at A* venues has four coauthors, and a paper published at A venues three more coauthors. This finding indicates that these papers require more collaboration (social capital) to examine various aspects of a problem, and thus require more knowledge (cultural capital) to be accepted for publication. Overall, female authors have more coauthors (85), while male authors have 78 fewer coauthors (8% throughout their career). Upon further examination of the distribution of coauthors, we discovered that 179 (4.4%) of the female authors collaborate primarily with females, while 13 020 (80%) of the male authors collaborate with male coauthors. 1758 (12%) of the male authors and 17 (0.4%) of the female authors had no publications with a coauthor of the other gender. On average, 3 (min: 0, max: 20, SD: 2, Mdn: 3) of the authors listed in a paper are male, demonstrating a disparity in collaboration between male and female authors.

The graph, illustrated in Fig. C1, comprises 34 components with 509 384 nodes (i.e. authors) and 1406 733 edges. The mean connectivity of a node is 5.9 (min: 1, max: 2108, SD:50), which means an author is somehow connected with around six other authors. The largest component contains 99.99% of the authors, indicating strong community connectivity. Based on the components, we identified 1212 941 cliques in the network. On average, a clique has a size of 3 (min: 2, max: 28, SD: 2), reflecting strong collaboration and the value of social capital in the S&P research community.

Evolution of gender authorship

Figure 12 provides an overview of the evolution of authorship by gender. Approximately 80% of the authors are male, and the number of female authors has only increased slightly over the decades (slope = 0.08). The fewest authors were observed in 1981, with a total of 25 (5 female and 20 male). The fewest female authors were observed in 1983, with only one female author. On average, female authors wrote 2 (min: 1, max: 111, SD: 4.8) papers and male authors wrote 3 (min: 1, max: 143, SD: 5.5). These numbers indicate that the S&P community is primarily male-dominated.

We performed a trend analysis of the numbers observed over the past 10 years to predict the number of female authors in the coming years. For this purpose, we used a linear regression model [70,71] to extrapolate the future presence of female authors. The model indicates that by 2029, the presence of female authors will increase by 29% compared to 2023. Despite the limited success of past efforts to increase the number of female authors, this trend may change in the future.

Female and male first authors

We identified 2716 (20%) papers with a female first author. Our data points show an increase in female first authors by 127% between 1980 and 2023. Despite the general stagnation in the number of female authors, these numbers suggest that a higher proportion of females are actively engaged in research activities within the field. Figure 13 illustrates the number of publications by female and male first authors. Overall, we identified 1687 (41%) distinct first female authors, compared to 6068 distinct male authors. 1242 (67%) of the female authors had one publication, and only 66 (4%) had more than five publications. The findings suggest

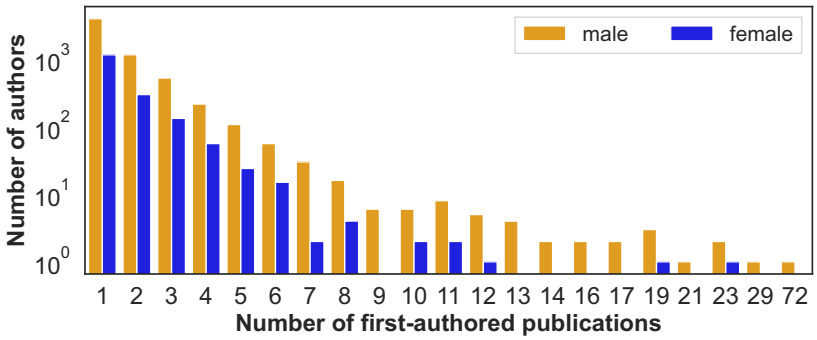

Figure 13 Number of authors who have made first-authored publications during their careers, divided by gender.

that female authors may be more likely to leave academia more frequently than their male counterparts, or at least, stop publishing more frequently [72].

Finally, we conducted a comparative analysis of the mean time intervals between consecutive publications. Our findings indicate that, on average, female first authors publish a paper every 1.1 (min 0:, max: 42, SD: 1.6) years, while male authors publish every 0.9 years (22% less). The median time between two publications is 2 years. 4,551 (24%) of the authors [male: 3669 (19%), female: 882 (5%)] have experienced a minimum of 3-year publication break during their academic career. The reasons for these breaks include parental leave, different obligations following the attainment of a doctorate or during the postdoctoral phase (e.g. teaching), or individuals leaving the academic field. It should be noted that if an author publishes in venues outside our scope, this information is not included in our data.

Female coauthor network

We have offered insights into the under-representation of female authors within the S&P community. In the following, we present further insights into the structure of female coauthor networks. We generated a subgraph of the coauthor network comprising only female authors to specifically analyze the evolution of female coauthor networks across the time spans defined in the section "Trends across venues." This analysis aims to understand how female authors are connected within the community (social capital), given that they publish less than their male colleagues (cultural capital). The results indicate that connectivity within the communities of female authors increased, demonstrating greater social capital. In subsequent periods, the number of these gatekeepers decreased. Currently, communities are characterized by a stronger meshwork of connections, with authors maintaining multiple connections within and outside their cliques. Additionally, authors often serve as "bridges," connecting disparate cliques of which they are part. Some authors also act as "liaisons," fostering connections between otherwise isolated cliques.

Discussion

The fields of scientometrics, metascience, and science of science are not the primary focus of the S&P community [73,74]. Only USENIX Security, among the top four ($\mathbb{A}^*$) conferences, listed metascience as a relevant topic in its 2025 call for papers but removed it in 2026. On the one hand, this observation explains why this important topic is not discussed in the S&P community; on the other hand, it poses a challenge for the community's further development, as it drives a focus on technical novelty without evaluating whether the community's efforts are fruitful. One prominent example of this is the popularity of research on *Blockchain Security and Privacy* in 2019, 2020, and 2021. Substantial parts of the community are working in this field. However, blockchain has never had a relevant impact outside the field of cryptocurrencies [75,76].

Applying Bourdieu's field theory, this pattern reflects the S&P field's doxa [77,78], i.e. the belief that legitimate security research is purely novel. Dominant agents, in our case, the program committees of the $\mathbb{A}$ and $\mathbb{A}^*$ S&P venues, exercise *symbolic violence* by excluding metascientific research and retrospective analysis as insufficiently novel. The blockchain case illustrates misrecognition. The field consecrated blockchain as important, and researchers invested heavily, believing this importance was objective. However, blockchain's prominence stemmed from symbolic struggles, such as competition for grants and prestige, rather than from demonstrated security impact. Without metascience to reveal these dynamics, the community cannot distinguish between grounded topics and elevated ones through "socially constructed enthusiasms" driven by external funding or media hype [79].

Evolution of S&P research

Over the past four decades, S&P research has expanded from basic security measures to diverse and complex challenges. Early work focused on access control, secure systems, and cryptography, but the field has since grown to include AI security, privacy regulations, and decentralized systems. While established topics like network intrusion detection remain relevant, new research domains continue to emerge, balancing foundational work with innovation. Major shifts in S&P research often align with external events and policy changes. High-profile cyber incidents have driven interest in threat intelligence and system resilience, while regulations like the EU GDPR have spurred privacy-focused research. These trends highlight a field that evolves in response to real-world challenges, continuously expanding its knowledge base to address an ever-changing threat landscape.

Recommendations

S&P research should move from a reactive approach to a more anticipatory approach by integrating strategic foresight into research planning. Long-term investment in underexplored areas is essential. Researchers should collaborate with policymakers and indus-

try leaders to shape regulations, ensuring that research informs policy rather than merely responding to it.

Collaboration in S&P research

Collaboration in S&P research strongly influences impact, with larger teams and broad coauthor networks leading to higher visibility and publication frequency. High-impact papers, especially in top-tier venues, often result from multidisciplinary teamwork, with A* publications typically having more coauthors than lower-tier ones. Although the community is well connected, dominant clusters centered on prolific research groups control much of the field's visibility. While these groups are innovation drivers, they may inadvertently limit access for junior or peripheral researchers.

Recommendations

To increase impact, the community should foster broader and more inclusive collaboration networks. Researchers should form interdisciplinary teams to approach problems from multiple angles, thereby increasing innovation and visibility. Established research groups should actively mentor and integrate early-career researchers, ensuring that collaboration networks are accessible rather than isolated. Universities and funding bodies should incentivize multi-institution projects and recognize team science in hiring and promotions, valuing collaborative contributions alongside individual achievements.

Diversity in the S&P community

S&P research remains male-dominated, with females comprising only 20% of authors, a ratio that has changed little over the decades. Female researchers have shorter academic careers, with many publishing only once before leaving academia. They tend to collaborate more, but remain underrepresented in high-impact networks. Additionally, they often specialize in specific subfields, which can limit visibility if those areas receive less attention. A persistent lack of diversity poses the risk of cognitive blind spots in S&P research, as homogenous teams may overlook critical threats or introduce biases. Addressing these disparities will strengthen the field by incorporating broader expertise and perspectives.

Recommendations

The S&P community must actively improve diversity and inclusion. Universities and conferences should implement outreach programs, scholarships, and recruitment efforts for underrepresented researchers, followed by structured mentorship and sponsorship initiatives. Encouraging mixed-gender collaborations is crucial, with project leaders integrating early-career female researchers into teams.

Lessons for research funding bodies

This study identifies structural dynamics within the S&P research community that are highly relevant to funding bodies aiming to support inclusive, impactful, and forward-looking science. Although the field has expanded significantly in research topics, vocabulary, and collaboration networks, this growth has not been accompanied by proportional diversification of its participant base. Female representation, in particular, has remained essentially unchanged in the field [80], but other forms of underrepresentation, such as regional, institutional, and epistemic diversity, also deserve attention. Furthermore, our findings reveal that the S&P community places high value on certain traditional academic forms of symbolic capital (e.g. top-tier venues, citation counts), which can reinforce existing hierarchies and limit greater participation. Finally, unlike disciplines such as medicine or psychology, S&P research rarely engages in self-reflective, data-driven evaluation of its structures and practices. This lack of metascientific engagement may hinder the community's ability to evolve transparently and equitably. Connecting these insights to broader policy frameworks, such as the EU's Horizon Europe diversity and research integrity goals, offers an opportunity to shape funding strategies that promote excellence and structural change.

Recommendations

Funding bodies should support targeted initiatives that address persistent gender imbalances and promote inclusive participation in S&P research. This includes investing in metascientific studies to improve transparency and self-evaluation within the field. Collaborative and interdisciplinary work should be actively encouraged, as it correlates with increased impact and knowledge exchange. Beyond traditional metrics such as citation counts, funders are encouraged to recognize diverse forms of academic contribution.

Conclusion

This study used a novel combination of two well-established sociological frameworks to evaluate the S&P research community along three dimensions. Specifically, we examined interactions among actors, venues, and outputs within the S&P research community and empirically investigated actors' accumulation of cultural, social, and symbolic capital over 44 years.

We conducted topic modeling on over 13k publications drawn from leading S&P venues between 1980 and 2023, identifying 100 research topics. We observe the growth of these research topics over time and provide insights into their emerging development, thus improving our understanding of S&P research output. The linguistic analysis reveals an increase in vocabulary fluctuation over the last decade and shifts in the co-occurrence of phrases. These observations indicate that the cultural capital of the S&P community has grown exponentially over time.

Our analysis reveals that female representation within the S&P community has remained largely unchanged over the past decades, suggesting the need for more inclusive growth. Our findings indicate a steady increase in the absolute and relative number of collaborations, suggesting that the community's social capital is developing and growing.

We provide insights into the factors that increase or decrease the impact of papers (i.e. citation counts). Our findings indicate that venue rank, number of authors, and abstract length significantly increase impact. As a result, the S&P community, like other research communities, strives to accumulate these scrutinized forms of capital.

Compared to other academic disciplines (e.g. medicine, physics, and psychology) [81,82], the S&P community underutilizes scientometrics and metascience, which can potentially impede the advancement of research quality. By improving our understanding of our community and highlighting areas that had not been previously explored, this study aimed to address this issue.

Author contributions

Nurullah Demir (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft, Writing–review & editing), Christian Böttger (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing–original draft, Writing–review & editing), Henry Hosseini (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing–original draft, Writing–review & editing), Meryem Demir (Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing–original draft), Christian Wressnegger (Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Validation, Writing–original draft), Norbert Pohlmann (Funding acquisition, Methodology, Project administration, Resources, Supervision), Tobias Urban (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft, Writing–review & editing).

Supplementary material

Supplementary material is available at *Journal of Cybersecurity* online.

Conflicts of interest

The authors have no competing interests to declare.

Funding

This work was supported by the German Federal Ministry of Education and Research (grant number UbiTrans 16KIS1629), the Federal Office for Information Security of Germany (grant number: 5Guide 01MO23033B), the Deutsche Forschungsgemeinschaft (German Research Foundation) (grant numbers 462287308 and BE1422/27-1), and the Helmholtz Association (HGF) (grant number “46.23 Engineering Secure Systems”), and Google’s educational cloud funding (grant number: EDU Credit 330580204).

Data availability

The data underlying this article, including the analysis code and data processing pipeline, are available in the repository at: https://pulse-of-cybersecurity.com/.

Appendix A: papers published by year

Figure A1 shows the number of papers published in each year on the analyzed venues. It is apparent that the community steadily drew over time, first rather slowly and then exponentially, demonstrating the demand for more secure and private technical solutions.

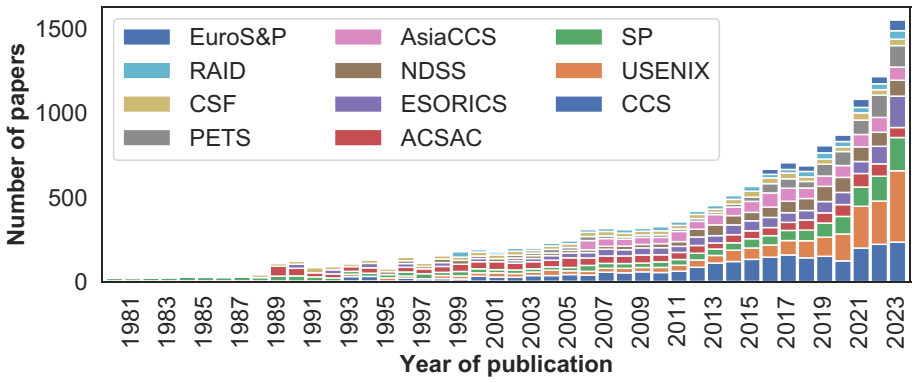

Figure A1 Number of papers by year and venue.

Appendix B: overview of identified topics

Table B1 provides a list of all 100 topics that we used in this work as a result of the clustering process (see the section "Data collection"). The topics are ordered by the number of papers in each.

Table B1 Identified topics in security and privacy research ordered by the number of papers in each topic.

ID	Topic name	∑	ID	Topic name
0	Processor side-channel attacks	469	51	Kernel rootkit detection and analysis
1	Android apps security and privacy	448	46	Cellular network privacy and security
4	Cryptographic security protocol analysis	423	67	Graph-based alert causality analysis
3	Blockchain security and privacy	414	49	Social network security and spam detection
9	Control flow integrity and code reuse attacks	361	76	SELinux and kernel integrity protection
2	Botnet detection and analysis	337	39	DNS security and privacy Measures
6	Access control security policies	320	52	Smart card security
5	Tor and network anonymity	313	54	Postquantum lattice-based cryptography
8	Intrusion detection systems	278	81	Identity-based encryption strategies
7	GDPR and data privacy practices	252	89	User awareness for security and privacy
13	Trusted computing and other trusted systems	249	72	Acoustic eavesdropping vulnerabilities
10	Adversarial attacks on neural networks	248	85	Machine Learning for Intrusion Detection
21	Secure signature schemes and protocols	236	42	Oblivious RAM for secure computation
11	Differential privacy mechanisms and guarantees	232	37	Internet censorship circumvention techniques
14	Secure multiparty computation techniques	231	62	Network vulnerability and attack strategy
18	Malware detection and deobfuscation techniques	230	45	Federated learning security techniques
15	Web security and browser vulnerabilities	226	68	Kernel fuzzing and driver Security
12	(Distributed) Denial of service attacks	218	50	Attribute-based encryption and access control
24	Secure information flow control	205	84	Patching and vulnerability management
17	Biometric security in digital devices	189	60	Forensic investigation and audit Log analysis
28	Information flow security analysis	182	58	Security of voice-controlled systems
20	Searchable encryption in databases	175	59	Security of Java runtime and applets
16	Location privacy	164	97	Authentication and key management protocols
25	Web tracking and browser fingerprinting	162	69	Facial recognition privacy and security risks
56	Kernel security and memory isolation	161	94	Sandbox security in untrusted code execution
19	IoT device security and privacy	153	96	Multifactor authentication strategies
27	Cloud storage security and verification	153	82	Homomorphic encryption techniques
29	Human factors in secure software development	146	64	Vehicle network security and robustness
22	TLS certificate security and validation	145	91	Binary code analysis techniques
47	Cryptographic security and analysis	145	63	Covert channels in networks
34	Membership inference attack strategies	143	80	Adversarial attacks in NLP
32	Fuzzing techniques	142	57	Genomic data privacy preservation
86	Buffer overflow vulnerability analysis	142	66	Bluetooth device security and privacy
31	Cryptographic implementation vulnerabilities and attacks	138	83	Email security and privacy enhancement
35	Trust management in PKI security	138	71	Secure deletion in flash storage
36	Web application vulnerability detection	134	75	Private information retrieval protocols
26	Verifiable electronic voting systems security	128	78	Cloud infrastructure security measures
30	Privacy-preserving machine learning techniques	126	90	Secure group communication protocols
43	Zero knowledge proof techniques	125	65	Secure private set intersection protocols
61	Authenticated key exchange protocols	124	73	Spam via SMS and phone calls
38	Security in web authentication and authorization	122	93	Privacy in social networks
48	Secure firewall and network configuration	122	98	Secure concurrency in database systems
23	Cyber physical systems and power grid security	117	79	SDN network security approaches
40	Password security and guessing attacks	117	95	Software supply chain security
55	Security in virtual machine hypervisors	113	74	RFID security and privacy

Table B1 Continued

ID	Topic name	∑	ID	Topic name	∑
41	Secure database management systems	111	88	Secure biometric authentication techniques	43
33	Website fingerprinting and traffic analysis	107	77	AI and machine learning in cybersecurity	40
44	Sensor network security and routing	96	99	Privacy protection in recommender systems	39
53	Firmware emulation and fuzzing security	94	87	Browser extension security and privacy	36
70	Wireless network security vulnerabilities	94	92	CAPTCHA security and countermeasures	34

Appendix C: plot of the coauthor network

Figure C1 depicts the generated co-author network of the analyzed papers. We build the graph using *Gephi* [83]. This graph demonstrates the interconnections among the authors, with nodes representing individuals and edges representing their collaborations. The figure indicates a strong connection among the authors of the S&P community.

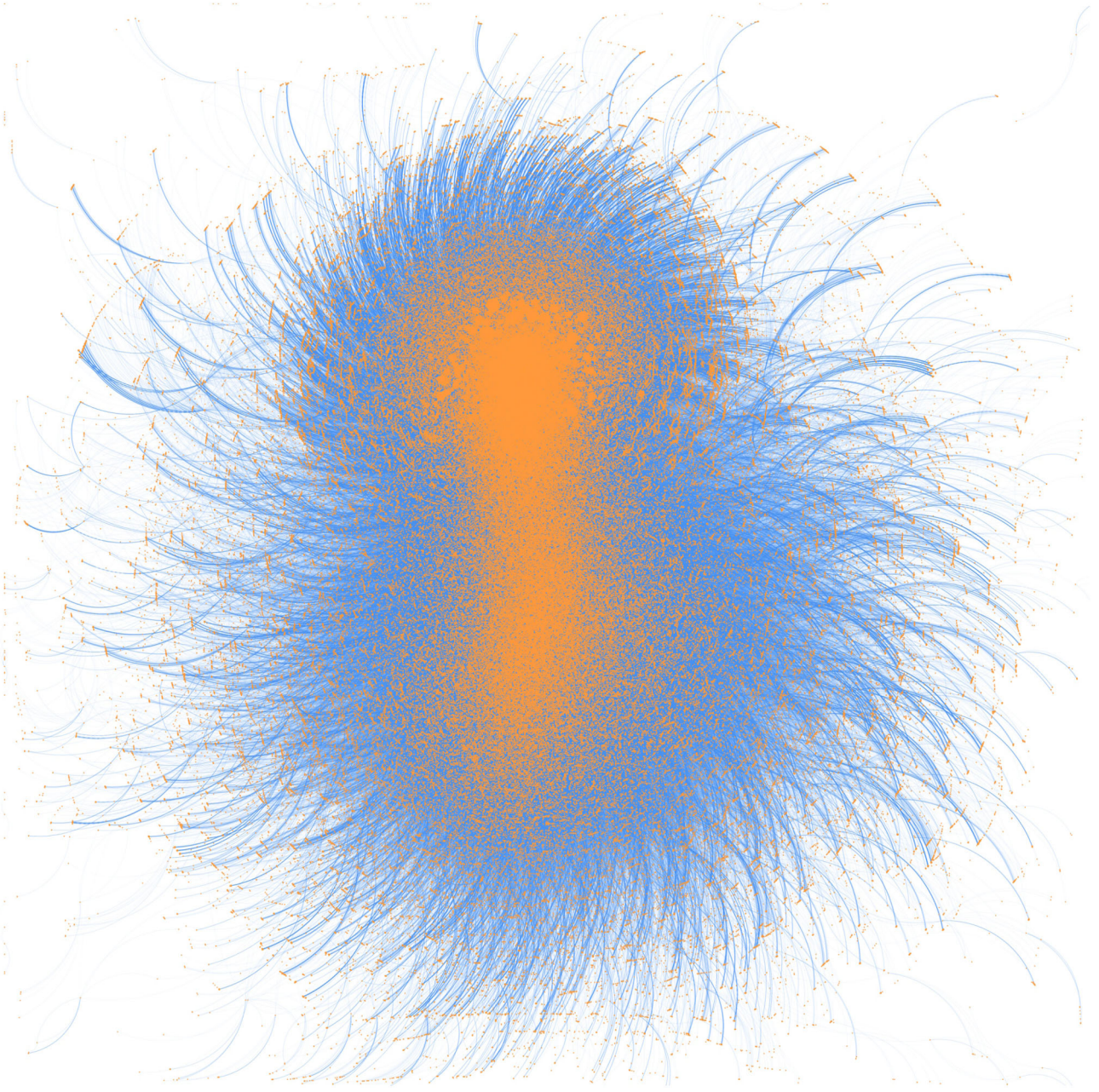

Figure C1 The plot illustrates the coauthor network from authors in our scope. Each node represents an author, and each edge represents a collaboration between two authors.

References

1. Balzarotti D. Welcome to the System Security Circus; 2026. https://www.s3.eurecom.fr/ balzarot/security-circus/ (16 March 2026, date last accessed).
2. Saltzer JH, Schroeder MD. The protection of information in computer systems. *Proc IEEE* 1975;**63**:1278–308. https://doi.org/10.1109/PROC.1975.9939.
3. Latour B. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press, 2005. https://doi.org/10.1093/oso/9780199256044.001.0001.
4. Bourdieu P. The forms of capital. In: JG Richardson (eds). *Handbook of Theory and Research for the Sociology of Education*. Westport: Greenwood, 1986, 241–58.
5. Grootendorst M. BERTopic: neural topic modeling with a class-based TF-IDF procedure. 2022. https://doi.org/10.48550/arXiv.2203.05794.
6. Martinez MA, Herrera M, López-Gijón J *et al.* H-Classics: characterizing the concept of citation classics through H-index. *Scientometrics* 2014;**98**:1971–83.
7. Leydesdorff L. The scientometrics challenge to science studies. *EASST Newslett* 1990;**9**:5–11.
8. Suryotrisongko H, Musashi Y. Review of cybersecurity research topics, taxonomy and challenges: interdisciplinary perspective. In:Unkown (ed.). *Proceedings of the IEEE 12th Conference on Service-Oriented Computing and Applications (SOCA)*. Piscataway: IEEE, 2019, 162–7. https://doi.org/10.1109/SOCA.2019.00031.

9. Baset A, Denning T. A {data-driven} reflection on 36 years of security and privacy research. In: *Proceedings of the 12th USENIX Workshop on Cyber Security Experimentation and Test*. New York: ACM, 2019.
10. S. M. Dhawan BMG, Elango B. Global cyber security research output (1998–2019): a scientometric analysis. *Sci Technol Lib* 2021;**40**:172–89. https://doi.org/10.1080/0194262X.2020.1840487.
11. Katsikeas S, Johnson P, Ekstedt M *et al.* Research communities in cyber security: a comprehensive literature review. *Comput Sci Rev* 2021;**42**:100431. https://doi.org/10.1016/j.cosrev.2021.100431.
12. Reuter C, Iacono LL, Benlian A. A quarter century of usable security and privacy research: transparency, tailorability, and the road ahead. *Behav Inf Technol* 2022;**41**:2035–48. https://doi.org/10.1080/0144929X.2022.2080908.
13. Loan FA, Bisma B, Nahida N. Global research productivity in cybersecurity: a scientometric study. *Global Knowl Mem Commun* 2022;**71**:342–54. https://doi.org/10.1108/GKMC-09-2020-0148.
14. Lee WH. How to identify emerging research fields using scientometrics: an example in the field of Information Security. *Scientometrics* 2008;**76**:503–25. https://doi.org/10.1007/s11192-007-1898-2.
15. Alqurashi F, Ahmad I. Scientometric analysis and knowledge mapping of cybersecurity. *Int J Adv Comput Sci Appl* 2024;**15**. https://doi.org/10.14569/IJACSA.2024.01503117.
16. Wendzel S, Lévy-Bencheton C, Caviglione L. Not all areas are equal: analysis of citations in information security research. *Scientometrics* 2020;**122**:267–86. https://doi.org/10.1007/s11192-019-03279-6.
17. Olijnyk NV. A quantitative examination of the intellectual profile and evolution of information security from 1965 to 2015. *Scientometrics* 2015;**105**:883–904. https://doi.org/10.1007/s11192-015-1708-1.
18. Lang M. The Fragmented Research Space of Cybersecurity: A Map of the Territory. In: *Proceedings of the European Interdisciplinary Cybersecurity Conference*. Cham: Springer, 2025, 116–132.
19. Koca M, Çiftçi S. A comprehensive bibliometric analysis of Big Data and Cyber Security: intellectual structure, trends, and global collaborations. *Knowl Inf Syst* 2025;**67**:1–26.
20. CORE Ranking. ICORE conference portal. 2026. https://portal.core.edu.au/conf-ranks/?search=4604&by=for&source=CORE2023&sort=arank&page=1 (16 March 2026, date last accessed).
21. Ley M. DBLP: some lessons learned. *Proc VLDB Endow* 2009;**2**:1493–500. https://doi.org/10.14778/1687553.1687577.
22. Crossref. Crossref: official website. 2026. https://www.crossref.org/ (16 March 2026, date last accessed).
23. Springer Nature. SpringerLink. 2026. https://link.springer.com/ (16 March 2026, date last accessed).
24. Institute of Electrical and Electronics Engineers. IEEE Xplore. 2026. https://ieeexplore.ieee.org/Xplore/home.jsp (16 March 2026, date last accessed).
25. Association for Computing Machinery. ACM Digital Library. 2026. https://dl.acm.org/ (16 March 2026, date last accessed).
26. Privacy Enhancing Technologies Symposium (PETS). PoPETS/PETS. 2026. https://petsymposium.org/ (16 March 2026, date last accessed).
27. Google LLC. Google Scholar. 2026. https://scholar.google.com/ (16 March 2026, date last accessed).
28. Cimpian JR, Kim TH, McDermott ZT. Understanding persistent gender gaps in STEM. *Science* 2020;**368**:1317–9. https://doi.org/10.1126/science.aba7377.
29. Master A, Meltzoff AN, Cheryan S. Gender stereotypes about interests start early and cause gender disparities in computer science and engineering. *Proc Natl Acad Sci* 2021;**118**:e2100030118. https://doi.org/10.1073/pnas.2100030118.
30. NamSor SAS. Namsor, the #1 AI for name origin, ethnicity & gender detection. 2026. https://namsor.app (16 March 2026, date last accessed).
31. Mayer N, Wendzel S, Keller J. Short paper: untersuchung des gender-gaps bei cybersecurity-publikationen (German). In: *Proceedings of the GI SICHERHEIT 2022*. Bonn: Gesellschaft für Informatik, 2022. https://doi.org/10.18420/SICHERHEIT2022_11.
32. Goyanes M, de Marcos L, Domínguez-Díaz A. Automatic gender detection: a methodological procedure and recommendations to computationally infer the gender from names with ChatGPT and gender APIs. *Scientometrics* 2024;**129**:6867–88. https://doi.org/10.1007/s11192-024-05149-2.
33. Sebo P, Shamsi A. Author gender and citation categorization: a study of high-impact medical journals. *Scientometrics* 2023;**128**:6299–306.
34. Ioannidis JPA, Boyack KW, Collins TA *et al.* Gender imbalances among top-cited scientists across scientific disciplines over time through the analysis of nearly 5.8 million authors. *PLoS Biol* 2023;**21**:e3002385. https://doi.org/10.1371/journal.pbio.3002385.
35. Squazzoni F, Bravo G, Grimaldo F *et al.* Gender gap in journal submissions and peer review during the first wave of the COVID-19 pandemic. A study on 2329 Elsevier journals. *PLoS One* 2021;**16**:1–17. https://doi.org/10.1371/journal.pone.0257919.
36. Löffler CS, Greitemeyer T. Are women the more empathetic gender? The effects of gender role expectations. *Curr Psychol* 2023;**42**:220–31. https://doi.org/10.1007/s12144-020-01260-8.
37. Fuentes A, Oyanadel C, Zimbardo P *et al.* Mindfulness and balanced time perspective: predictive model of psychological well-being and gender differences in college students. *Eur J Invest Health Psychol Educ* 2022;**12**:306–18. https://doi.org/10.3390/ejihpe12030022.
38. Santamaría L, Mihaljević H. Comparison and benchmark of name-to-gender inference services. *PeerJ Comput Sci* 2018;**4**:e156. https://doi.org/10.7717/peerj-cs.156.
39. Sebo P. Performance of gender detection tools: a comparative study of name-to-gender inference services. *J Med Lib Assoc* 2021;**109**:414–21.
40. Benz P, Pradier C, Kozlowski D *et al.* Mapping the unseen in practice: comparing latent Dirichlet allocation and BERTopic for navigating topic spaces. *Scientometrics* 2025;**130**:1–32.
41. Ma L, Chen R, Ge W *et al.* AI-powered topic modeling: comparing LDA and BERTopic in analyzing opioid-related cardiovascular risks in women. *Exp Biol Med* 2025;**250**:10389.

42. Blei DM, Lafferty JD. Dynamic topic models. In: *Proceedings of the 23rd International Conference on Machine Learning. ICML '06*. New York: ACM, 2006, 113–20. https://doi.org/10.1145/1143844.1143859.
43. Blei DM, Lafferty JD. Topic models. In:Unknown (eds). *Text mining*. London: Chapman and Hall/CRC, 2009, 101–24.
44. NLTK Project. NLTK :: Natural Language Toolkit. 2026. https://www.nltk.org/ (16 March 2026, date last accessed).
45. Řehůřek R, Sojka P. Software framework for topic modelling with large corpora. In: Unkwnon (eds). *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Valletta: ELRA, 2010, 45–50. http://is.muni.cz/publication/884893/en (16 March 2026, date last accessed).
46. Newman D, Lau JH, Grieser K *et al.* Automatic evaluation of topic coherence. In: Unkwnon (eds). *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. New York: ACM, 2010, 100–8.
47. Aletras N, Stevenson M. Evaluating topic coherence using distributional semantics. In: Unknown (eds). Association for Computational Linguistics Potsdam, Germany *Proceedings of the International Conference on Computational Semantics*. 2013.
48. Mimno D, Wallach HM, Talley E *et al.* Optimizing semantic coherence in topic models. In: R Barzilay, M Johnson (eds). *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. EMNLP '11*. Edinburgh: Association for Computational Linguistics, 2011, 262–72.
49. Hoyle A, Goel P, Peskov D *et al.* Is automated topic model evaluation broken? The incoherence of coherence. In: M Ranzato, A Beygelzimer, Y Dauphin, PS Liang, JW Vaughan (eds). *Advances in Neural Information Processing Systems. Vol. 34 of NIPS '21*. Red Hook: Curran Associates, Inc., 2021, 2018–33.
50. Reimers Nils, Gurevych Iryna. Pretrained models–sentence transformers documentation. 2026. https://www.sbert.net/docs/sentence_transformer/pretrained_models.html (16 March 2026, date last accessed).
51. Grootendorst M. Embedding models. 2026. https://maartengr.github.io/BERTopic/getting_started/embeddings/embeddings.html. (16 March 2026, date last accessed).
52. Ermakova L, Bordignon F, Turenne N *et al.* Is the abstract a mere teaser? Evaluating generosity of article abstracts in the environmental sciences. *Front Res Metrics Anal* 2018;**3**:16.
53. Demir N, Große-Kampmann M, Urban T *et al.* Reproducibility and replicability of web measurement studies. In: *Proceedings of the ACM Web Conference 2022*. New York: Association for Computing Machinery, 2022, 533–44. https://doi.org/10.1145/3485447.3512214.
54. Grootendorst Maarten. LLM & Generative AI–BERTopic. 2026. https://maartengr.github.io/BERTopic/getting_started/representation/llm.html#chatgpt. (16 March 2026, date last accessed).
55. Rayson P, Garside R. Comparing corpora using frequency profiling. In:Unknown (eds). *The Workshop on Comparing Corpora*. Hong Kong: Association for Computational Linguistics, 2000, 1–6.
56. Wilson A. Embracing Bayes factors for key item analysis in corpus linguistics. In: M Bieswanger, A Koll-Stobbe (eds). *New Approaches to the Study of Linguistic Variability*. Lausanne: Peter Lang, 2013, 3–11.
57. Rychlý P. A lexicographer-friendly association score. In:Unknown (eds). *Proceedings of the 2nd Workshop on Recent Advances in Slavonic Natural Language Processing. RASLAN 2008*. Brno: Masaryk University, 2008, 6–9.
58. Amos R, Acar G, Lucherini E *et al.* Privacy policies over time: curation and analysis of a million-document dataset. In:Unknown (eds). *The Web Conference 2021–Proceedings of the World Wide Web Conference. WWW '21*. New York, ACM, 2021, 2165–76. https://doi.org/10.1145/3442381.3450048.
59. Hosseini H, Utz C, Degeling M *et al.* A bilingual longitudinal analysis of privacy policies measuring the impacts of the GDPR and the CCPA/CPRA. *Proc Priv Enhancing Technol* 2024;**2024**:434–63. https://doi.org/10.56553/popets-2024-0058.
60. Rescorla E. *The Transport Layer Security (TLS) Protocol Version 1.3*. RFC Editor, 2018. RFC 8446. https://www.rfc-editor.org/info/rfc8446 (16 March 2026, date last accessed). https://doi.org/10.17487/RFC8446.
61. COCCIA M. How does science advance? Theories of the evolution of science. *J Econ Social Thought* 2020;**7**:153–80. https://doi.org/10.1453/jest.v7i3.2111.
62. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc* 1952;**47**:583–621.
63. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J* 1948;**27**:379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.
64. Rosseel Y. Lavaan: an R package for structural equation modeling. *J Stat Softw* 2012;**48**:1–36. https://doi.org/10.18637/jss.v048.i02.
65. Ullman JB, Bentler PM. *Structural Equation Modeling, Chapter 23*. Hoboken: John Wiley & Sons, Ltd, 2012. https://doi.org/10.1002/9781118133880.hop202023.
66. Lu J, Velten B, Klaus B *et al.* Meta-research: the changing career paths of PhDs and postdocs trained at EMBL. *eLife* 2023;**12**:e78706. https://doi.org/10.7554/eLife.78706.
67. Liu P. Computer security conference acceptance rates. GitHub, 2026. https://github.com/puzhuoliu/Computer-Security-Conference-Acceptance-Rate. (16 March 2026, date last accessed).
68. European University Association Council for Doctoral Education. *Doctoral Education in Europe Today: Approaches and Institutional Structures*. Brussels: European University Association, 2019. (7 November 2024, date last accessed).
69. Zweben S, Bizot B. 2021 Taulbee survey: doctoral degree production recovers from the pandemic; undergraduate enrollment remains high but shows first indications of a slowdown. *Comput Res News* 2022;**34**:1–48.
70. James G, Witten D, Hastie T *et al. An Introduction to Statistical Learning: with Applications in R*. Vol. **103**. New York: Springer, 2013.
71. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Vol. **2**. New York: Springer, 2009.
72. Mason MA, Goulden M. Do babies matter (Part II)? Closing the baby gap. *Academe* 2004;**90**:10–15.
73. Le Pochat V, Joosen W. Analyzing cyber security research practices through a meta-research framework. In: *Proceedings of the 16th Cyber Security Experimentation and Test*

Workshop. CSET '23. New York: Association for Computing Machinery, 2023, 64–74. https://doi.org/10.1145/3607505.3607523.
74. Demir N. Better Web Measurements: Enhancing Security and Privacy Research. Karlsruhe: Karlsruher Institut für Technologie (KIT), 2023. https://doi.org/10.5445/IR/1000163503.
75. AlShamsi M, Al-Emran M, Shaalan K. A systematic review on Blockchain adoption. *Appl Sci* 2022;**12**. https://doi.org/10.3390/app12094245.
76. Bracci E, Tallaki M, Ievoli R *et al.* Knowledge, diffusion and interest in blockchain technology in SMEs. *J Knowl Manage* 2021;**26**:1386–407. https://doi.org/10.1108/JKM-02-2021-0099.
77. Bourdieu P. *Outline of a Theory of Practice. Cambridge Studies in Social and Cultural Anthropology*. Cambridge: Cambridge University Press, 1977.
78. Bourdieu P, Thompson JB. *Language and Symbolic Power*. Cambridge: Harvard University Press, 1991.
79. Bourdieu P. The specificity of the scientific field and the social conditions of the progress of reason. *Social Sci Inf* 1975;**14**:19–47. https://doi.org/10.1177/053901847501400602.
80. Garr-Schultz A, Muragishi GA, Mortejo TA *et al.* Masculine defaults in academic science, technology, engineering, and mathematics (STEM) fields. *Psychol Sci Public Interest* 2023;**24**:1–9.
81. Ioannidis JP, Fanelli D, Dunne DD *et al.* Meta-research: evaluation and improvement of research methods and practices. *PLoS Biol* 2015;**13**:1–7. https://doi.org/10.1371/journal.pbio.1002264.
82. Ioannidis JPA. Meta-research: why research on research matters. *PLoS Biol* 2018;**16**:1–6. https://doi.org/10.1371/journal.pbio.2005468.
83. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In:Unknown (eds). *Proceedings of the international AAAI conference on web and social media*. Vol. **3**. Washington: AAAI Press, 2009, 361–2.

Received: 22 May 2025. **Revised:** 19 January 2026. **Accepted:** 18 February 2026