Security Dynamics in Computer Science Technologies

March 2022

Dimitri Percia David[a,b,*]; William Lacube[b]; Sébastien Gillard[c]; Thomas Maillart[a]; Alain Mermoud[b,**]; Loïc Maréchal[d]; Michael Tsesmelis[b]

[a] Information Science Institute, Geneva School of Economics and Management, University of Geneva, 40 Boulevard du Pont-d'Arve, 1211 Geneva, Switzerland.
[b] Cyber-Defence Campus, armasuisse Science and Technology, Feuerwerkstrasse 39, 3602 Thun, Switzerland.
[c] Chair of Defense Economics, Military Academy at ETH Zurich, Kaserne Reppischtal, 8903 Birmensdorf, Switzerland.
[d] Faculty of Economics and Business, University of Neuchatel, Rue Abraham-Louis-Breguet 2, 2000 Neuchâtel, Switzerland.

[*] Corresponding author: `dimitri.perciadavid@unige.ch`; CYD Campus, EPFL Innovation Park, building I (3rd floor), 1015 Lausanne, Switzerland.

[**] Permanent address: `alain.mermoud@ar.admin.ch`; CYD Campus, EPFL Innovation Park, building I (3rd floor), 1015 Lausanne, Switzerland.

Declaration of competing interest: This research received a Cyber-Defence (CYD) Campus Fellowship, through the EPFL Research Office.

For readability purposes, the use of colors is needed for all figures of this manuscript.

Highlights

Security Dynamics in Computer Science Technologies

- Quantitative framework for the assessment of *security dynamics* in technologies.
- Identification of 3 patterns among 20 computer-science technology categories:
- 1. Technological and security developments are not correlated.
- 2. Security gains more attention at a later stage of technological development.
- 3. *Opinion* on technology is associated with security development.

Security Dynamics in Computer Science Technologies

ARTICLE INFO

Keywords:
technology mining
technology management
security economics
time-series econometrics
natural language processing
opinion mining

ABSTRACT

The quantitative study of *security dynamics* in computer-science technologies is essential for understanding security-development patterns of information systems. Here, we specify and investigate *security dynamics* as (i) the relation between technological and security developments, (ii) the security development, modeled as the evolution of *security considerations* among technologies, and (iii) the effect of security development on the *opinion* given to technologies. We perform a scientometric analysis on `arXiv` e-prints ($n = 340\,569$) related to 20 computer-science technology categories. Our empirical results are threefold. First, we provide evidence of a lack of relation between the technological and security developments: while most categories follow a sigmoid-growth curve of technological development, this latter is not a determinant of security development. Second, we find a *security-attention* pattern: over the lifetime of categories, *security considerations* appear more frequently, emphasizing that security gains more attention at a later stage of technological development. Third, we find an *opinion* pattern: the experts' *opinion* related to each category is positively explained by the prevalence of *security considerations*. These results emphasize new methods for understanding, modeling, and benchmarking *security dynamics* of technologies, which brings new heuristics for considering changes related to the security of information systems.

1. Introduction

The accelerating pace of technological development is continuously redefining information and communication technologies (ICTs) [83]. Emerging technologies present a myriad of opportunities to enhance the efficacy and efficiency of operations for all types of social organizations [15]. Yet, in such a fast-paced and complex context of technological development, opportunities are undeniably also accompanied by threats [52, 8, 47].

Counter-measures to contain information-security threats have been widely investigated, developed, and implemented. For instance, *secure-by-design* (SBD) engineering implies considering security as the first stage of technological development [7]. Yet, research in security economics demonstrates that security is often ill-developed in the early stages of product development [16]. The misaligned incentives between end-users and information-security producers and providers, the high investment that information security requires and its engineering complexity hinder security development among technologies [8, 16, 7].[1]

While factors limiting security development have been the subject of a growing body of literature, there is dearth of knowledge regarding the actual security development among technologies and its relation with technological development, which could shed light on how security evolves within systems. This is a central question when considering the evolution footprint of security among technologies, a notion we define as *security dynamics*. We therefore investigate three *security dynamics* of interest: (i) the relationship between technological and security developments, (ii) the evolution of the attention towards *security considerations* embedded in those technologies, and (iii) the effect of *security considerations* on the *opinion* towards technologies.[2] We adopt a technology-mining approach – *i.e.*, scientometrics and its related methods – on 1 854 076 e-prints of the `arXiv` open-data repository (from August 14, 1991, until December 31, 2020). Out of this sample, we identify 340 569 e-prints related to 20 *Computer-Science Technology Categories related to Cybersecurity* (hereafter, CSTCCs), on which we investigate three *security dynamics* mentioned above. To the best of our knowledge, this work presents the first indicator for capturing security dynamics in computer-science technologies.

Our results are threefold. First, we find empirical evidence of a lack of relationship between technological development and security development. While most CSTCCs follow a sigmoid growth pattern of technological

ORCID(s):

[1] By *security development*, we mean the development of *information security*. This latter is the practice of protecting information through the dependability of technology (in terms of privacy-preserving and confidentiality aspects, as well as the ability of technology to ensure the integrity, availability, and non-repudiation of data). Information security typically operates through the mitigation of information risks that involve probability prevention or reduction of unauthorized/inappropriate access of data, unlawful use, disclosure, disruption, deletion, corruption, modification, inspection, recording, or devaluation of information [7].

[2] We define *security considerations* as employed semantics related to *information security*.

development, their security development does not follow such a pattern. Moreover, the change rates of these two patterns are uncorrelated. Second, we find empirical evidence of a *security-attention* pattern. For most CSTCCs, the prevalence of *security considerations* grows with time, supporting the hypothesis that security is taken into account to a greater extent at a later stage of technological development. Third, we find empirical evidence of an *opinion* pattern. The prevalence of *security considerations* is positively correlated to the *opinion*, supporting the hypothesis that CSTCCs with a greater prevalence of *security considerations* trigger a more favorable *opinion*. Also, its standard deviation – the *opinion* dispersion – is negatively related to the *opinion*, supporting the hypothesis that *consensus* is associated with a more favorable *opinion*.

These results emphasize unfolded central aspects related to the structural dynamics of security development among computer-science technologies. Moreover, such elements constitute relevant and informative considerations for developing security among technologies. We discuss how such enlightening considerations may be considered for shaping social change in security-development guidelines and principles. More specifically, we suggest using the quantitative evaluation performed in this work to create informative security-development benchmarks across the technological development field and thus target specific security development according to weak links (*i.e.*, technologies which perform poorly) in the domain.

The remainder of this article proceeds as follows. Section 2 grounds this research with a critical literature review. Section 3 presents the theoretical framework and related hypotheses. Section 4 details the data and the methodology. Section 5 presents the results. Section 6 discusses the limitations and sets a future work agenda. Section 7 serves as the conclusion.

2. Related work

In this section, we review the methods developed for measuring the theoretical variables related to (i) technological development, (ii) security development, and (iii) the relation between *opinion* mining and security. We emphasize the research gaps that we exploit in this work through such a review process, giving relevancy to our approach. Such gaps are the scarcity of (i) benchmarking indicators related to technological development within an inter-technologies context, (ii) holistic indicators of security development, and (iii) how *security considerations* are associated with *opinion*.

2.1. Assessment of technological development

Technological development (also called *technological change*) is the overall invention, innovation, and diffusion of technologies [46]. Technological development undergoes the first stage of engineering origination (*i.e.*, invention) of one or various features of a technology, the second stage of practical implementation (*i.e.*, innovation) of them, and the last stage of commercialization or release (*i.e.*, diffusion) of such technology throughout the market [80].

Numerous approaches, methods, and models have been developed to identify, assess, and forecast the aforementioned stages of technological development (for an extensive literature review, see [26, 53, 39, 17, 29, 75]). Among these approaches, a central one consists of modeling technological development as a general trend pattern that follows an *S-curve* – *i.e.*, a sigmoid curve – which depicts the development performance of technology through time [57, 79, 80, 60, 22, 2, 5]. Such a model describes a *life-cycle* of technological development through three phases: an introduction phase, a growth phase, and a maturity phase [76].

At the introduction phase of a technology, development performance changes occur relatively slowly. Such a pace depends on the specific technology and its environment. However, it is not uncommon to witness years of gestation before seeing an emerging technology achieving widespread acceptance and commercial success. By satisfying the users' needs of niche market segments, a novel technology improves efficacy and efficiency before reaching a broader population of mainstream users [3]. Then, during the growth phase, the technology is exposed to the mass market. As a result, the technology becomes more compelling, attracting more investment and incremental innovation (new features). Such a growth phase – characterized by the migration of the technology from niche markets to mainstream markets – leads to a swift proliferation of new features of the technology and more companies involved with the technology [50, 72]. Lastly, as technology begins to mature, the pace of technological development typically slows down as it approaches its performance limits and diffusion limits due to market share saturation [73, 1].

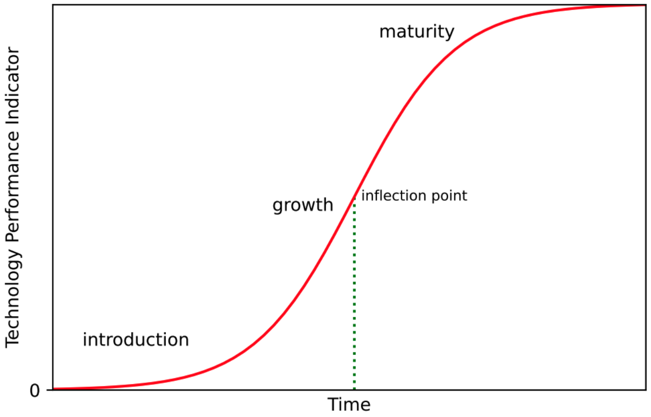

Figure 1: ***S-curve* of technological development**. The y-axis can be interpreted as a *technology-performance indicator*, in which, for instance, the diffusion process of a technology can be captured. Within the growth phase, an inflection point (projected into the x-axis, in dashed green) is reached, and corresponds to the moment where the growth rate of technological development is diminishing, though still positive.

In practical terms, such an S-curve can be measured through various private- and open-source data indicators. Past research has relied mainly on indicators such as bibliometric analysis related to scientific publications (*i.e.*, scientometrics) [31, 45, 77], patents [21, 35, 69], industry-market indicators (job openings, trade registers), or a mix of them [27, 54]. In this aspect, text-mining methods [21, 24, 41] and network analytics [57, 64] are particularly prolific. For instance, concerning text-mining methods, Guo et al. (2019) analyze 1 666 scientific publications to detect trends and hot-spots (*i.e.*, hypes) in network technologies and information systems [37]. With a literature-growth approach and a co-citation analysis, they find exponential patterns – capturing the introduction and the initial growth phases of the S-curve – in technological diffusion and recurring themes related to specific categories. In another example, Priestley et al. (2020) trace the growth curve of Web technologies between 1990 and 2013. By investigating a longitudinal dataset of 20 493 Web-related US patents, they find that the accumulation of corporate Web inventions follows an S-shaped curve [76]. Son et al. (2010) assess global trends in automation and robotics technologies concerning network-analytics methods [84]. They identify main research topics and link them to the primary contributors by analyzing peer-reviewed publications.

The great majority of these approaches have mainly been employed for specific technologies [29]. Yet, such specific analysis impedes a broader investigation of a shared S-curve pattern of technological development within an ensemble of technologies. To the best of our knowledge, no study has quantitatively investigated such a development pattern among various technologies for comparison purposes, at least not for *Computer-Science Technology Categories related to Cybersecurity* (CSTCCs). Consequently, there is a need to create a benchmark for technologies by a shared development metric, as the absence of such a metric thwarts the technological development comparison among different CSTCCs.

2.2. Assessment of security development

Assessing security-development indicators among computer-science technologies has been done through creation and analysis of cybersecurity skills indexes (*e.g.*, [18]), organizational development of cybersecurity programs (*e.g.*, [36]) and job openings (*e.g.*, [10]), cybersecurity risks (*e.g.*, [44]), dynamics of cybersecurity incidents (*e.g.*, [59]), evolution of cybersecurity behaviors (*e.g.*, [56]), and so on. For a systematic literature review, see [63].

However, while quantitative methods are widely applied to technological development analysis, to the best of our knowledge, the quantitative investigation of a holistic, dynamic, and common metric for measuring development patters of security among technologies still constitutes an unexplored domain in the scientific literature. Consequently, there is a need to model a general security-development indicator to assess how *security considerations* evolve within technologies.

2.3. Assessment of *opinion* mining given *security considerations*

Trend analysis is a central aspect of technology assessment and forecasting. It is used to capture development patterns and thus to gauge the attention (*i.e.*, the *hype*) that a community is giving to a technology [28]. However, trends do not capture the *opinion* of this given community, as the above-mentioned qualitative approaches and their methods are, in essence, ignoring the experts' *opinion* on a given technology. [49]. Yet, such experts' *opinion* is an interesting indicator for evaluating to what extent technology is believed to be relevant, effective, and efficient in a specific context [58]. As authors of academic works related to technology are mainly made up of engineers and scholars who evaluate – and more than not also help develop – this same technology, they can be considered experts. Therefore, the semantics they use to describe the technology can be analyzed and aggregated to classify their *opinion* towards this same technology, complementing trend analysis. Capturing *opinion* is possible with quantitative methods such as *sentiment analysis* [58].

While capturing the experts' *opinion* on a given technology complements trend-analysis methods, investigating which factors may influence such an *opinion* could bring interesting insights. For instance, the extent to which experts emphasize *security considerations* attached to technology may influence their *opinion* about this same technology [74]. Yet, to the best of our knowledge, the extent to which the experts' *opinion* on a given technology may be correlated to the prevalence of *security considerations* they raise on this same technology has not been investigated. However, such an investigation is relevant as it can shed light on how security is associated with the perception of a technology – which is assuredly a critical aspect.

3. Theoretical framework and hypotheses

In this section, we ground our hypotheses related to the potential patterns of *security dynamics* that we want to test for the ensemble of *Computer-Science Technology Categories related to Cybersecurity* (CSTCCs). We divide such patterns into three categories: (i) the relationship dynamics between technological and security developments, (ii) the dynamics of *security attention*, and (iii) the dynamics between *opinion* and *security considerations*.

3.1. Technological and security-development patterns

Before investigating the relationship between technological and security developments, we first need to model the first one. As depicted in Section 2, the sigmoid function exhibits an initial exponential rate of technological diffusion, then reaches an inflection point and settles into a phase of diminishing returns to scale, and finally saturates [80] – reaching technological maturity (see Figure 1). Similar technological development evidence has been investigated in various fields (*e.g.*, [22, 13, 60]). Additionally, Rogers (2010) states that technological diffusion occurs within a social system [80]. In this respect, the `arXiv` community is a typical social system in which e-prints related to technologies are communicated through uploads on a shared repository. Furthermore, in many technical fields such as mathematics, physics, and computer science, an important share of research papers are self-archived on the `arXiv` repository before being submitted and subsequently presented or published either at a conference or in a peer-reviewed journal [86]. Therefore, we consider `arXiv` to be an adequate, relevant, and representative source for measuring the development patterns through the count of scientific works per month related to a technical subject. Consequently, the monthly number of uploads of e-prints related to CSTCCs can be used to assert whether technological development does indeed follow a sigmoid growth pattern. Therefore, Hypothesis 1a:

H1a: *The technological development of each CSTCC follows a sigmoid pattern.*

Even if we observe idiosyncratic growth rates for each CSTCC, CSTCCs likely follow a shared sigmoid-like development pattern. However, despite a theoretical background stating such claims, we are not aware of any quantitative work investigating the presence of such a shared technological development pattern among CSTCCs. Moreover, even if **H1a** is not directly related to *security dynamics*, testing it is a necessary first stage for testing **H1b**, which is directly related to *security dynamics*.

The technological development of computer-science technologies encompasses a broad spectrum of engineering and project-management steps, going from the analysis of novel usability needs to technical development up to the very last deployment aspects of the technology [92]. In this whole technological development life-cycle, technical aspects related to *security considerations* are evaluated, designed, and implemented [42]. These considerations are taken into account – although within largely variable magnitudes, depending on the technological development context [7].

However, numerous scholars, policy-makers, and practitioners emphasized discrepancies among drivers of technological development on the one hand and security development on the other hand. For instance, researchers in information-security economics highlighted that misaligned incentives between developers and end-users constitute a significant barrier in the security development of technologies (*e.g.*, [8, 7, 6, 9, 16]). Notably, Anderson & Moore

(2006) stated that incentives matter at least as much as technical aspects for the security development of large-scale systems. In other words, the security failures of technologies arise when individuals who could fix them are not the ones who suffer the costs of such failures [8]. Consequently, such a misalignment of incentives may tend to detach the security development of CSTCCs to their technological life-cycle dynamics.

For instance, Anderson (2007) pointed out that software markets have some of the characteristics of a *market for lemons* [9]. Following the concept of Akerlof (1970) [4], the author examines how the quality of information-security products in the software market is degraded in the presence of information asymmetry between the buyers and the sellers concerning the security efficacy of those same products. Should it be for security software, or the security of software developed for other purposes, most users cannot to tell what is vulnerable (and which technology is dependable) and how efficient (and dependable) is the information-security product in solving such a vulnerability issue. Therefore, the buyers' willingness-to-pay will shrink. Consequently, sellers won't be incentivized to produce better quality for less money – *i.e.*, developers are not compensated for efforts to strengthen their code. Such a *market for lemons* mechanism may then detach security development from other technological development aspects [9].

Similarly, market dynamics such as the need to secure business gaps – by launching innovative products on the market as soon as possible – tend to put security-development aspects at a less considered matter. In other words, sellers often forego *security considerations* by offering underdeveloped security technologies and then using clients as continual beta testers for identifying and patching vulnerabilities [7]. Moreover, security aspects are often skipped in the early stages of technological development as substantial investments are required to tackle such security aspects, consequently decreasing profitability [8, 16]. Therefore, Hypothesis 1b:

H1b: *The security development of a CSTCC is uncorrelated with its technological development.*

In other words, the security development of CSTCCs likely follows a distinctive path compared to the technological development path. However, despite a theoretical background stating such claims, there is no quantitative work investigating the differences – and most importantly, the correlation – between technological and security development among CSTCCs.

3.2. *Security-attention* pattern

In addition to investigating a potential lack of correlation between technological development and security development, we argue that out of the bibliometric measure of *security considerations* depicted in e-prints – and previously used as a proxy of security development –, a measure of *security attention* can be derived. Based on the dynamics of the mean of *security considerations*, such a measure can shed light on the development of the attention given to security among each CSTCC.

Analyzing the mean's trend of *security considerations* is relevant when evaluating the extent to which the CSTCC experts (*i.e.*, the authors of e-prints) pay attention to *security considerations* in their fields of expertise. The evolution of such a measure constitutes an informative indicator to gauge the dynamics of *security considerations* raised among each CSTCC. The example of the *security by design* (SBD) engineering principle and its actual implementation constitutes a relevant illustration for understanding what we mean by measuring the dynamics of *security attention*. Information-security threats have brought fundamental engineering-perspective changes in the field of computer-science technologies, such as developing technologies according to the SBD principle [19, 51]. Under SBD, the security development of software and information systems is embedded within the technological development process [81]. Products, services, and capabilities are thus designed from the very beginning to be secure [19]. However, as stated above, a product must be put on the market quickly to become profitable as fast as possible, and such attitudes thus relegate security issues to a later phase of development [16]. These works indirectly suggest that *security considerations* would grow only at a later stage of technological development. Therefore, Hypothesis 2:

H2: *The* security attention *of each CSTCC increases over time.*

Despite a theoretical background stating that economic factors hinder the SBD implementation [8, 7, 16], there is no quantitative works that measure the dynamics of *security attention*. Therefore, testing this hypothesis will shed light on the development of *security considerations* related to CSTCCs.

3.3. *Opinion* pattern

Like investigating a potential *security-attention* pattern, assessing a possible *opinion* pattern brings a complementary analysis to technological development.

Sentiment analysis (also called *opinion mining*) refers to the study of individual sentiment, typically captured by the analysis of the employed semantics of different text sources and translated into computer-readable metrics with the help of natural language processing (NLP). Researchers have widely used the method to analyze the sentiment of a community towards a variety of topics from marketing, management, or finance – to name a few [58]. A central assumption of opinion mining is that sentiment is the product of the translation of individuals' judgment, thinking, and attitudes towards a given topic [34, 20]. Hence, observing the sentiment constitutes a measure of the aforementioned individual peculiarities (see [61]). Empirically speaking, the sentiment is also often used to predict market trends [12]. In the context of technology assessment, capturing the sentiment that experts – *i.e.*, authors of e-prints – develop towards a particular CSTCC might give insights into where their *opinion* towards the CSTCC currently stands. Even though it is commonly admitted that researchers might be opinion-neutral in their work, they still use natural language to describe their findings on technologies. Thus, they cannot escape the unavoidable tendency to express words and ideas that are either partial or not towards the technology.

Interestingly, the sentiment relates to the prevalence of *security considerations*. Gurung & Raja (2016) show that the prevalence of privacy and security aspects provided on a given topic effects individuals' perceived risk concerning this exact topic [38]. Yet, the connection between risk factors and sentiment is widely studied in multiple disciplines (*e.g.*, [20]). By capturing social consensus through sentiment analysis, Yang et al. (2016) showed that consensus and perceived risk are related [89]. Moreover, Yang et al. (2015) show that technological uncertainty – captured by sentiment analysis – determines (among other factors) the perceived risk [90]. Consequently, perceived risks, captured by technological uncertainty and security concerns, are supposedly linked to a measure of sentiment. Therefore, the prevalence of *security considerations* emitted by experts on a given CSTCC should be positively related to the sentiment (*i.e.*, *opinion*) that such experts have towards this same CSTCC. Therefore, Hypothesis 3a:

H3a: *For each CSTCC, the experts'* opinion *is positively related to its prevalence of* security considerations.

Additionally, the iterative nature of both product engineering (through design thinking) and peer-reviewing of scientific works (through design science) brings to the fore an incremental process helping the development of products and innovations [85]. Such an incremental process eventually leads to a consensus on a given topic. For example, Dou et al. (2017) and Lehrer & Wagner (2012) show that the *opinion* related to this exact product tends to converge towards a consensus through a given product evolution [32, 55]. Similarly, Yüzügüllü and Deason (2007) show that the technical maturity and market-readiness of technology are factors that facilitate the consensus of the community on these same technologies [91]. Such convergence in *opinion* (*i.e.*, a consensus) reflects a decreasing dispersion (a decreasing standard deviation) of this same *opinion* [43]. Hence, we expect to uncover similar evidence in the technology field in general and in scientific works related to technology development in particular. Consequently, the experts' *opinion* on a given CSTCC might be negatively associated with the extent to which such an *opinion* is polarized (depicting a greater dispersion). In other words, the consensus (inversely related to the *opinion* dispersion) that experts have on a given technology might be positively associated with the *opinion* emitted by these same experts on this same technology. Therefore, Hypothesis 3b:

H3b: *For each CSTCC, the experts'* opinion *is positively related to its* consensus.

Testing these two hypotheses allows to uncover and disentangle two elements of the experts' *opinion* emitted on each CSTCC: first by understanding the link between the prevalence of *security considerations* and *opinion*, and second by highlighting the positive correlation between this *opinion* and the *consensus* within the community of experts.

4. Data and methods

In this section, we first present the empirical variables and how we measure them. These empirical variables – namely the (i) e-prints, the (ii) *security considerations*, and the (iii) sentiment – are the ones we use to capture our constructs – namely, the (i) technological development, (ii) security development, (iii) *security attention*, and (iv) *opinion*. We then present the methodologies we use to test our hypotheses related to each construct.

4.1. Data

We extract data from open scientific works (*i.e.*, scholar articles consisting of working papers, preprints, technical reports, post-proceedings, and publications) labeled *e-prints* and uploaded on the `arXiv` repository. This latter is a free distribution service and open-access archive for academic articles related to various technical fields, including computer science (uploaded e-prints are not peer-reviewed).[3] First, we download the entire `arXiv` repository (1 858 293 files, corresponding to 3 TB of text in *.pdf* format) through a mirror of the database found on `kaggle`.[4] The data encompasses all e-prints uploaded since the inception of the `arXiv` repository (August 14, 1991) until December 31, 2020. Next, for each *Computer-Science Technology Categories related to Cybersecurity* (CSTCC), we (i) count the number of e-prints through time, (ii) assess the prevalence of *security considerations* present in these e-prints, and (iii) assess the *opinion* expressed by authors.

4.1.1. e-prints

The empirical variable *e-prints* is the main variable as it encompasses (i) the statistics of scientific work uploads related to each CSTCC – used as a proxy of technological development –, (ii) the text used for capturing *security considerations* – a construct that is also subsequently used for capturing *security attention* –, and the *opinion*.

To consistently classify and archive all e-prints, `arXiv` representatives – composed of a scientific advisory board – created a systematic category taxonomy.[5] They determined this taxonomy with a Delphi-like method involving expert members for each `arXiv` scientific field.[6] This implies that authors who want to upload their e-prints on `arXiv` must select the corresponding category. Then, `arXiv` moderators check the authors' classification to ensure consistency. We consider this 3-step classification to be robust as (i) the taxonomy is created through a consensus reached by a panel of experts, (ii) authors have no apparent incentive to misclassify their work, and (iii) moderators check the classification consistency. As e-prints are attached to various predetermined `arXiv` fields unrelated to computer science (such as physics, mathematics, quantitative biology, quantitative finance, and economics), we filter the `arXiv` predetermined fields to extract computer-science technologies, namely the *computer science* (denoted `cs.`) repository. We apply a second filter, considering `arXiv` subcategories in the `cs.` fields that are directly associated with information-security technologies. To determine which `arXiv` subcategories of the `cs.` repository are effectively related to information-security technologies, we use the *Defenses* sections listed in the *Information Security* portal of *Wikipedia* as a reference.[7] Thus, we select the `arXiv` subcategories of the `cs.` repository whenever this subcategory is also mentioned within the Wikipedia *Defenses* section.

From this 2-step selection procedure, 20 subcategories are retained. In the case of the `cs.` repository, the category taxonomy substantially relies on the list of methodology and technology categories provided by the *2012 ACM Computing Classification System*.[8] Therefore, we consider the 20 categories mentioned above as distinct CSTCCs. We depict the list of these CSTCCs and their respective number of e-prints in Table 1 on page 8.

If the `arXiv` repository is nowadays regarded as an established platform amongst various scientific communities for uploading their e-prints, it enjoyed no such popularity at its inception. Therefore, we cannot assume that the `arXiv` platform depicts a constant attention rate related to each CSTCC. To circumvent this bias, we normalize the number of e-prints related to each CSTCC by dividing the total number of e-prints per CSTCC per period (month) by the corresponding amount of total e-prints (*i.e.*, including all categories) of the `arXiv` repository per period (month). Such a measure is depicted in Figures 2 and 3 on page 8. A preliminary analysis shows that, for the great majority of categories, we either witness an exponential trend – depicted in Figure 2 (*i.e.*, corresponding to the *introduction* and the *growth* phases of the *S-curve*) – or a proper sigmoid trend – depicted in Figure 3 (*i.e.*, corresponding to the three phases of the *S-curve*).

[3] https://arxiv.org/

[4] https://www.kaggle.com/Cornell-University/arxiv

[5] https://arxiv.org/about/people/scientific_ad_board#advisory_committees

[6] https://arxiv.org/category_taxonomy

[7] https://en.wikipedia.org/wiki/Information_security

[8] https://arxiv.org/corr/subjectclasses

`arXiv` categories	Cluster name (CSTCC)	Total count of e-prints	With *security considerations*	% of *security considerations*
`cs.AI`	Artificial Intelligence	38 620	11 447	29.640
`cs.AR`	Hardware Architecture	2573	971	37.738
`cs.CC`	Computational Complexity	8492	1216	14.319
`cs.CL`	Computation and Language	29 528	8536	28.908
`cs.CR`	Cryptography and Security	19 784	14 952	75.576
`cs.CV`	Computer Vision and Pattern Recognition	64 696	21 852	33.776
`cs.DB`	Databases	6269	2341	37.342
`cs.DC`	Distributed, Parallel, and Cluster Computing	14 955	5686	38.021
`cs.DS`	Data Structures and Algorithms	18 269	3458	18.928
`cs.GT`	Computer Science and Game Theory	7992	2279	28.516
`cs.HC`	Human-Computer Interaction	8774	2753	31.377
`cs.IR`	Information Retrieval	10 407	3216	30.902
`cs.LG`	Machine Learning	94 024	30 142	32.058
`cs.NE`	Neural and Evolutionary Computing	10 155	2649	26.086
`cs.NI`	Networking and Internet Architecture	16 606	6826	41.106
`cs.OS`	Operating Systems	652	303	46.472
`cs.PL`	Programming Languages	5731	1937	33.799
`cs.RO`	Robotics	16 187	6055	37.407
`cs.SE`	Software Engineering	10 032	4109	40.959
`cs.SY`	Systems and Control	18 347	6845	37.309

Table 1: **`arXiv` categories (corresponding CSTCCs) and their respective count of e-prints, with and without *security considerations***. Not surprisingly, the category `cs.CR` (*Cryptography and Security*) has a share of *security considerations* greater than 75%, which validates our method for capturing *security considerations*; see subsection 4.1.2 on page 9. As almost all categories are serially correlated, descriptive statistics of the normalized number of e-prints are not presented here (should be interpreted with care). Such statistics are available upon request.

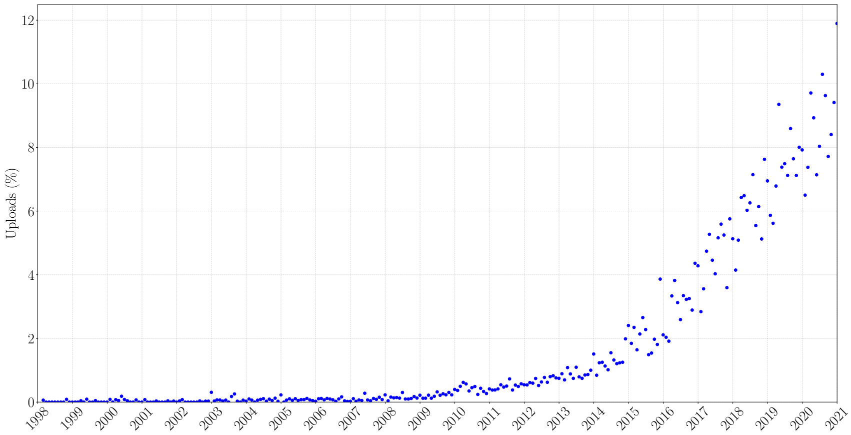

Figure 2: **Normalized count of e-prints: computer vision and pattern recognition**. The frequency is monthly. The plot clearly pictures the *introduction* and *growth* phases of the *S-curve* (see Figure 1 on page 3).

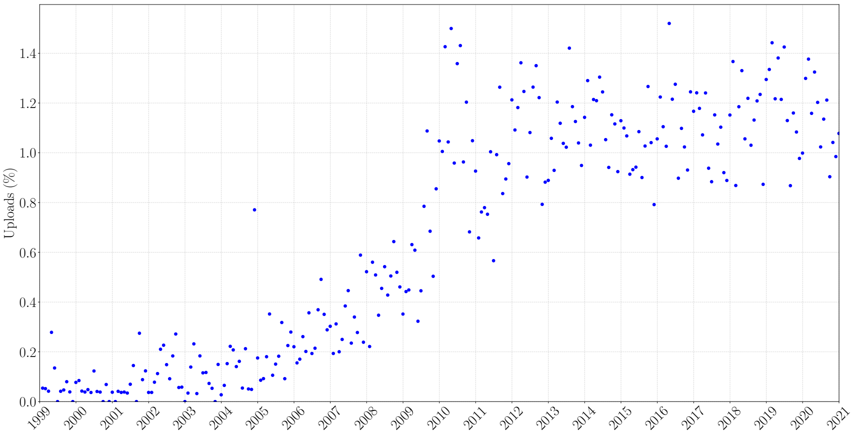

Figure 3: **Normalized count of e-prints: networking and internet architecture**. The frequency is monthly. The plot clearly pictures the three phases of the *S-curve*, namely the *introduction*, *growth* and *maturity* phases (see Figure 1 on page 3).

4.1.2. Security considerations

As previously defined, our construct named *security considerations* is related to (i) the technology's dependability in terms of privacy-preserving and confidentiality aspects, and (ii) the technology's ability to ensure the integrity, availability, and non-repudiation of data. To capture *security considerations* expressed in e-prints, we thus select a set of keywords related to these concepts mentioned above. These relate to the well-known CIA triad (*i.e.*, *confidentiality*, *integrity*, and *availability*), and the *non-repudiation* principle [23, 78] and are depicted in the *Information Security* portal of *Wikipedia*.[9] The list of keywords is: *secure*, *security*, *safe*, *reliability*, *dependability*, *confidential*, *confidentiality*, *integrity*, *availability*, *defense*, *defence*, *defensive*, and *privacy*.

We then query the `arXiv API` to select e-prints that contain these keywords in either their title or abstract. Subsequently, to extract the prevalence of *security considerations* among each CSTCC, we divide the number of e-prints per CSTCC including these keywords, by the total number of e-prints per CSTCC. Figure 4 depicts how the share of e-prints alluding to security has changed, illustrated by the CSTCC *computer vision and pattern recognition*. A preliminary analysis shows that, for the great majority of categories, we witness (i) a diminishing dispersion and (ii) an upward trend of the measure.

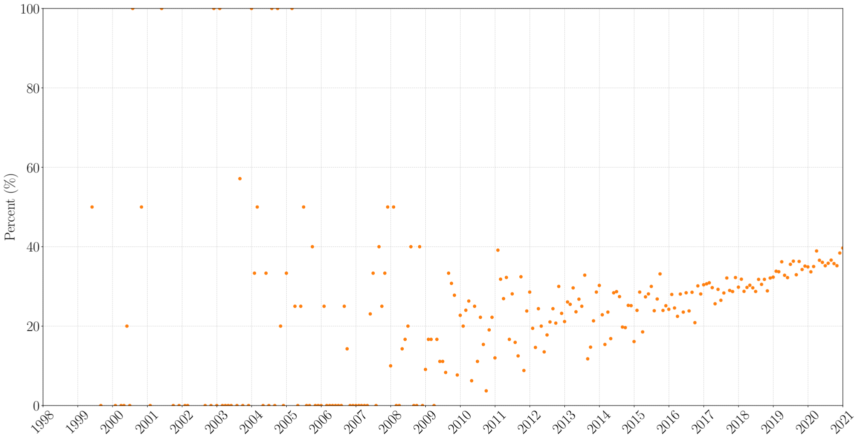

Figure 4: ***Security considerations*: computer vision and pattern recognition**. This figure depicts the evolution of the prevalence of *security considerations*, *i.e.*, the e-prints containing *security considerations* divided by the total number of e-prints. The frequency is monthly. Due to the *Law of Large Numbers*, the left-hand side of the plot – corresponding to more sparse data – does not show interesting properties. However, the right-hand side depicts a diminishing dispersion, and – more importantly – an upward trend. Such a pattern is witnessed in the great majority of categories, as depicted in the multi-plot of all *security considerations* measures (for each CSTCC) available in Figure 11 on page 23 (Appendix). As for e-prints, almost all categories are serially correlated. Therefore, descriptive statistics of *security considerations* are not presented here as they are to be interpreted with care. Such statistics are however available upon request.

4.1.3. Sentiment

To capture the *opinion* of authors related to each e-print attached to a given CSTCC, we employ sentiment analysis by implementing a classical lexicon-based approach based on a labeled thesaurus (the NLTK *sentiment lexicon* of `Python`, in English natural language) to classify the semantics of authors as either positive or negative [82].

To do so, we first clean and normalize every word in e-prints before transforming them into tokens (*i.e.*, machine-readable inputs). Cleaned tokens are obtained through standard NLP procedures such as (i) transforming all text in GB English, (ii) removing special characters, *stop words*, punctuation, and lowering upper-cases. Then, we normalize cleaned tokens through lemmatization (morphological analysis to transform tokens into their canonical form).

Subsequently, we apply a standard cumulative-sentiment function that classifies each token into either a positive or negative sentiment before summing the result for each e-print. The final sentiment score occupies a range from -1 (for the worst sentiment) to 1 (for the best sentiment), giving a normalized sentiment score for each e-print. The scores of e-prints related to the same CSTCC are then cumulated for each month (for the publication date and not the upload date). Statistics depicting the sentiment distribution are then available for each CSTCC and for each month. Figure 5 shows an example of the evolution of the sentiment for the CSTCC *computer vision and pattern recognition*. Descriptive statistics of the sentiment for each CSTCC are listed in Table 2.

[9] `https://en.wikipedia.org/wiki/Information_security#Key_concepts`

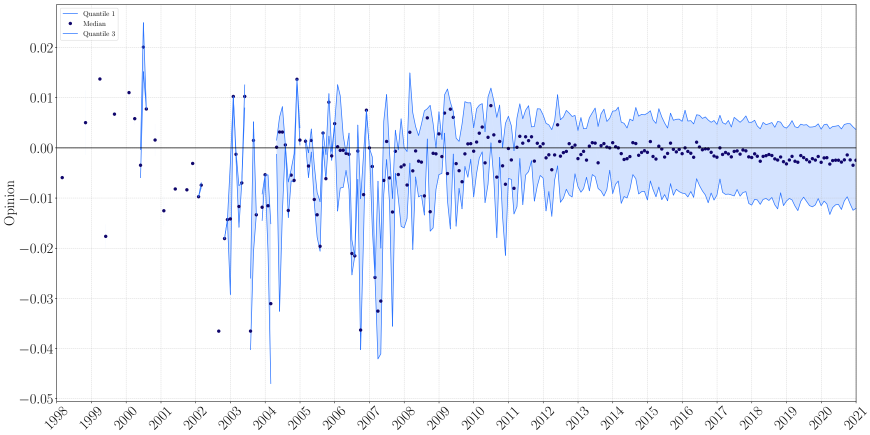

Figure 5: **Distribution of *opinion*: computer vision and pattern recognition**. The median is plotted with dots and the second and third quartiles are plotted with lines. The frequency is monthly. Similarly to Figure 4, the plot shows no interesting properties on his left-hand side as the data are sparse. However, a decreasing dispersion (due to the *Law of Large Numbers*) – and more importantly – a downward trend is perceptible. A multi-plot of all *opinion* measures (for each CSTCC) available Figure 11 on page 23 (Appendix).

`arXiv` categories	**Mean**	**Median**	**Std Dev**	**Skewness**	**Kurtosis**
`cs.AI`	−0.002	−0.001	0.006	−1.341	2.809
`cs.AR`	−0.001	0.000	0.007	−1.592	7.664
`cs.CC`	−0.009	−0.009	0.005	−0.967	3.658
`cs.CL`	0.004	0.005	0.003	−1.038	3.462
`cs.CR`	−0.006	−0.005	0.014	−10.379	140.804
`cs.CV`	−0.003	−0.002	0.007	−2.053	7.548
`cs.DB`	0.001	0.001	0.006	−0.077	9.936
`cs.DC`	−0.001	0.000	0.005	−1.570	11.395
`cs.DS`	−0.004	−0.003	0.005	−0.008	9.114
`cs.GT`	0.002	0.003	0.008	−1.379	14.532
`cs.HC`	0.004	0.005	0.007	−1.261	5.772
`cs.IR`	0.006	0.007	0.007	−3.640	22.239
`cs.LG`	−0.002	−0.001	0.006	−1.737	10.354
`cs.NE`	−0.003	−0.001	0.007	−2.538	13.560
`cs.NI`	−0.002	−0.001	0.005	−1.774	13.112
`cs.OS`	−0.003	−0.001	0.010	−1.441	4.284
`cs.PL`	0.002	0.002	0.006	−3.476	40.685
`cs.RO`	−0.003	−0.002	0.007	−3.337	18.244
`cs.SE`	−0.001	0.000	0.006	−1.022	5.650
`cs.SY`	−0.005	−0.005	0.004	−0.529	7.064

Table 2: **Descriptive statistics: monthly *opinion***. This table displays summary statistics of the monthly *opinion* for each CSTCC.

4.2. Methods

The following subsection presents the methodologies employed to test our hypotheses. For all methods, we define a set Ω_x for all CSTCC, x:

$$\Omega_x = \left\{ t \mid t \leq N_x, t \in \mathbb{N}^* \right\} \tag{1}$$

where N_x is the number of months comprised between the first and the last e-print for x.

4.2.1. Technological and security-development patterns

To model the **technological development**, we apply a *Levenberg-Marquardt* algorithm of non-linear optimization for fitting a noiseless sigmoid curve on each CSTCC development (*i.e.*, a clean technological development trend – based on the *S-curve* theory – to be fitted to the historical data points of e-prints) [65]. We use the `Python scipy` package and its `.optimize.curve_fit` method to fit a sigmoid function into each CSTCC sample.[10] We made this choice because the sigmoid function precisely depicts S-shaped curve characteristics: it is a bounded, differentiable, and real function defined for all real input values and has both a non-negative derivative at each point and precisely one inflection point [40]. Therefore, we define the technological development function, $\sigma_x(t)$, with $t \in \Omega_x$,

$$\sigma_x(t) = \frac{L_x}{1 + e^{-k_x(t - t_{0x})}} \tag{2}$$

where:

- t_{0x} is when the inflection point is reached (corresponding to maximum of the first derivative of the function, *i.e.*, the maximum growth rate of technology development [80]);
- L_x is the curve's maximum limit value (*i.e.*, $\lim_{t \to +\infty} \sigma_x(t) = L_x$) [88, 80];
- k_x is the sigmoid growth rate or steepness of the curve [88, 80].

For every time-series of e-prints related to a CSTCC, D_x, the `.optimize.curve_fit` method finds the optimal values of the parameters L_x, k_x and t_{0x} and their standard errors (by minimizing non-linear least-squares errors).[11] If fitting such a sigmoid function to our datasets yields compelling metrics – *i.e.*, (i) if 90% of data-points fall within the boundaries of the standard error of the regression, and (ii) if the goodness-of-fit statistic, the reduced chi-squared, $\chi^2_{\nu_x} \approxeq 1$ –, then **H1a** would be verified for a given x.[12]

We proceed to a multivariate time-series analysis for each x concerning the **security development** and its relationship with technological development. More precisely, we fit a multivariate autoregressive model to each x, which comprises both autoregressors (of order $p_x \in \Omega_x$) of the dependent variable – *i.e.*, the security development, S_x –, and regressors (of order $q_x \in \Omega_x$) of the independent variable – *i.e.*, the technological development, D_x. Both orders are determined by using the selection process of the *Akaike info criterion* [11]. In other words, we investigate if the security development of a given CSTCC is explained by the past values of its technological development. We model such a multivariate autoregressive model to each x as follows,

$$S_{x,t} = \zeta_x + \sum_{i=1}^{p_x} \phi_{x,i} S_{x,t-i} + \sum_{j=1}^{q_x} \theta_{x,j} D_{x,t-j} + u_{x,t} \tag{3}$$

where:

- S_x is the time series of security development, $S_{x,i}$ is its value at time i, and $\phi_{x,i}$ is its autoregressive parameter;[13]
- D_x is the time series of technological development, $D_{x,j}$ is its value at time j, and $\theta_{x,j}$ is its regressor parameter;
- ζ_x is a constant;
- $u_{x,t}$ is the time-series of error terms (*i.e.*, $S_{x,i} - \hat{S}_{x,i}$).

[10] The *.optimize.curvefit* method is a method for minimizing objective functions, possibly subject to constraints. It includes solvers for non-linear problems (with support for both local and global optimization algorithms), linear programming, constrained and non-linear least-squares, root finding, and curve fitting (see, `https://docs.scipy.org/doc/scipy/reference/optimize.html`).

[11] *Non-linear least squares* is the form of least squares analysis used to fit a set of v observations with a model that is non-linear in w unknown parameters ($v \geq w$). The basis of the method is to approximate the model by a linear one and refine the parameters by successive iterations.

[12] In our case, the standard error of the regression is captured by computing the squared root of the reduced chi squared, denoted as χ^2_ν. The χ^2_ν statistic, also known as the mean squared weighted deviation (MSWD), is used as a goodness-of-fit metric. This statistic can be interpreted as follows: a $\chi^2_\nu \gg 1$ indicates a poor model fit. A $\chi^2_\nu > 1$ indicates that the fit has not fully captured the data (or that the error variance has been underestimated). In principle, a value of χ^2_ν around 1 indicates that the extent of the match between observations and estimations is in accord with the error variance. A $\chi^2_\nu < 1$ indicates that the model is overfitting the data: either improperly fitting noise or overestimating the error variance [14].

[13] NB: As we capture security development through the dynamics of *security considerations*, here, S_x is used interchangeably for both security development and *security considerations*.

Such multivariate autoregressive models are estimated through ordinary least squares (OLS). As time-series S_x and D_x have their respective seasonality for each x, we proceed to a seasonal adjustment by using the *Seasonal and Trend decomposition using Loess* (STL) method [25]. Also, we proceed to a logarithmic transformation of the above-mentioned series as they often present exponential growth [11]. Finally, as these time-series are trended, we need to stationarize them through respective differentiation of order $I_{D_x}(n)$ and $I_{S_x}(m)$, where n and $m \in \mathbb{N}^*$ [11].

If the model presents compelling performance metrics – *i.e.*, *adjusted* R^2 –, and at least one statistically significant and positive estimator for a regressor of the independent variable (*i.e.*, the technological development), then **H1b** would be verified.[14]

4.2.2. Security-attention pattern

To capture the *security-attention* pattern, we compute a rolling mean, Γ, of *security considerations* to capture the mean's trend of such a variable. In other words, for each CSTCC, we compute the annual mean trend of the share of e-prints that encompass security aspects. We model the rolling mean for each x as follows,

$$\Gamma_{S_{x,t}} = \frac{1}{12} \sum_{i=t-11}^{t} S_{x,i} \tag{4}$$

Suppose such a rolling mean of the prevalence of *security considerations* displays an increasingly positive trend. In that case, it will support an increase in the prevalence of *security considerations* over time. Thus, **H2** would be verified.

4.2.3. Opinion pattern

We analyze the determinants of the experts' *opinion* – *i.e.*, the author's *opinion* – by examining the effect of two variables on this same *opinion*. The first is the variable *security considerations* (**H3a**), while the second is the standard deviation of the *opinion* as a measure of *opinion* dispersion (**H3b**).

To test **H3a** and **H3b**, we use the cross-sectional approach of Fama & MacBeth [33]. This method, originally developed to estimate both market-risk exposures and risk premia of assets, is a two-pass estimation. We use the second pass, a sequence of cross-sectional OLS regressions at each month t, with $t \in \Omega_x$, for an x of the form,

$$y_x = \alpha_x + \beta_x \eta_x + \gamma_{\mathbf{x}} \mathbf{Z_x} + \epsilon_x \tag{5}$$

where:

- y_x is a measure of opinion;
- α_x is a constant;
- η_x is the variable of interest (*i.e.*, the prevalence of *security considerations*), and β_x is its estimator;
- $\mathbf{Z_x}$ is a matrix of additional controls, and γ_x is a vector of estimators;
- and ϵ_x is the error term.

Next, we consider the estimated time series of β, $\hat{\beta}$, to test whether they significantly depart from zero. In addition, we correct for serial correlation and heteroskedasticity in $\hat{\beta}$ with Newey-West's adjustment method [67].

We project the time-series of parameters on a constant and extract the covariance matrix of errors that we adjust to retrieve the standard errors. As the procedure of [67] implies specifying ahead the number of lags, we also use the non-parametric approach of [68] with automatic lag selection. Despite the small size of the cross-section (20 CSTCCs), the large time-series dimension still permits statistical inference.[15] Finally, we consider the unavailability of some CSTCCs at the beginning of our sample and restrict the estimation to a period starting in November 2002, when the cross-section reaches 15 simultaneous observations for the first time (for a total of 217 time-observations in our sub-sample). To proxy for the instantaneous *opinion*, we consider, in turn, the median and mean of the aggregated *opinion*. Finally, to control that our results are not driven by the numerator or denominator of the prevalence of *security considerations*, we control the number of e-prints with *security considerations* and the total number of e-prints, respectively.

[14] Before interpreting the results, we verified that the time-series of $u_{x,t}$ is not (i) serially correlated and (ii) is not heteroskedastic [11]. Such statistics are available upon request.

[15] In fact, in their study, Fama and MacBeth (1973) use a cross-section of only 20 portfolios.

5. Results

In this section, we present the results of applying specified methods (Section 4) we use to test our hypotheses (Section 3). **H1b**, **H2**, **H3a**, and **H3b** are verified for all *Computer-Science Technology Categories related to Cybersecurity* (CSTCCs), while **H1a** is verified for the great majority of CSTCCs. Our *Python* script is available upon request.

5.1. A sigmoid trend as a technological development pattern

Table 3 shows the metrics and parameter values of non-linear regressions that fit a sigmoid function to our data.

arXiv categories	χ^2_ν	SE	L	k	t_0
cs.AI	10.062	3.172	49 908.902	0.015	2071
cs.AR	1.889	1.375	1297.227	0.016	2065
cs.CC	2.588	1.609	0.489	0.032	2004
cs.CL	12.145	3.485	5.568	0.039	2019
cs.CR	3.013	1.736	10.783	0.015	2028
cs.CV	4.966	2.228	13.626	0.037	2019
cs.DB	2.454	1.567	0.461	0.025	2010
cs.DC	2.178	1.476	1.665	0.020	2015
cs.DS	2.788	1.670	1.273	0.037	2009
cs.GT	1.969	1.403	0.524	0.054	2009
cs.HC	2.275	1.508	1766.652	0.020	2051
cs.IR	2.137	1.462	8.690	0.015	2031
cs.LG	12.952	3.599	12 463.478	0.030	2039
cs.NE	3.042	1.744	1.298	0.025	2016
cs.NI	2.383	1.544	1.125	0.046	2008
cs.OS	0.893	0.945	78.830	0.007	2115
cs.PL	2.712	1.647	0.423	0.022	2011
cs.RO	3.762	1.940	6.918	0.032	2022
cs.SE	3.297	1.816	0.855	0.025	2013
cs.SY	10.444	3.232	2.963	0.031	2018

Table 3: **Sigmoid fits of monthly normalized and aggregated number of e-prints per CSTCC**. This table displays the goodness-of-fit measures (*i.e.*, the χ^2_ν, and the regression standard error (SE)), and parameters of the sigmoid fits of the total normalized e-prints per CSTCC. The parameter t_0 indicates the year in which the maximum growth rate of the CSTCC is reached.

The results are compelling: out of the 20 CSTCCs, 15 exhibit $\chi^2_\nu > 1$, 5 exhibit $\chi^2_\nu \gg 1$, and 1 exhibits $\chi^2_\nu < 1$.[16] Therefore, the sigmoid function fits our observed data for 15 different CSTCCs. The unsatisfactory results (*i.e.*, fits that yield a $\chi^2_\nu < 1$ and a $\chi^2_\nu \gg 1$) can be explained as follows. The only fit that yields a $\chi^2_\nu < 1$ is cs.OS. Such a poor fit is results from a high dispersion (important standard deviation) of the cs.OS data; a high dispersion due to the sparsity of the data (in fact, only 652 e-prints have been uploaded for the whole history of the arXiv repository). Concerning the fits that yield a $\chi^2_\nu \gg 1$, we systematically witness the start of the sigmoid function of technology development without witnessing the inflection point yet. This means that, for some cases, the *Levenberg-Marquardt* algorithm struggles to calibrate (optimize) the parameters. CSTCCs concerned by this issue are (i) cs.AI, (ii) cs.CL, (iii) cs.LG, and (iv) cs.SY. Figure 6 on page 14 shows the fit of a typical sigmoid growth pattern onto the CSTCC cs.DS, while Figure 7 on page 14 shows the fit of a typical exponential growth pattern onto the CSTCC cs.CV; this exponential rise can be regarded as the left side (*i.e.*, beginning) of a sigmoid growth pattern. Figure 8 on page 14 normalizes all above-mentioned fits for comparison purposes, while Figure 12 on page 24 plots the sigmoid fits for all CSTCCs. Hence, **H1a** is verified for at least 15 CSTCCs, which exhibit a common development pattern that follows a sigmoid function. Concerning the above-mentioned CSTCCs that yield a $\chi^2_\nu \gg 1$, we still witness an exponential growth pattern that coincides with the beginning of a sigmoid function. If we cannot confirm that these five CSTCCs will automatically follow a sigmoid technology-development growth, we cannot reject this hypothesis either. Such results support the theory that most CSTCCs follow a typical *S-curve* pattern of technology development [80].

[16] *i.e.*, the different χ^2_ν lay within the same order of magnitude than 1, and are greater than one for 15 CTSCCs out of 20.

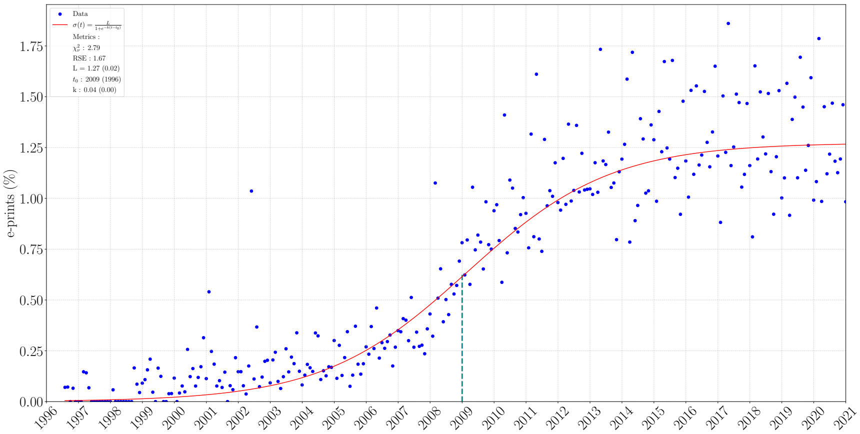

Figure 6: **Sigmoid fit of e-prints: data structures and algorithms**. The normalized e-prints are in blue, and the sigmoid fit (equation 2) is in in red. We additionally plot the inflexion point (vertical green dashed segment). We report the parameters of the sigmoid fit, their standard errors in parenthesis and the χ_ν^2. The frequency is monthly.

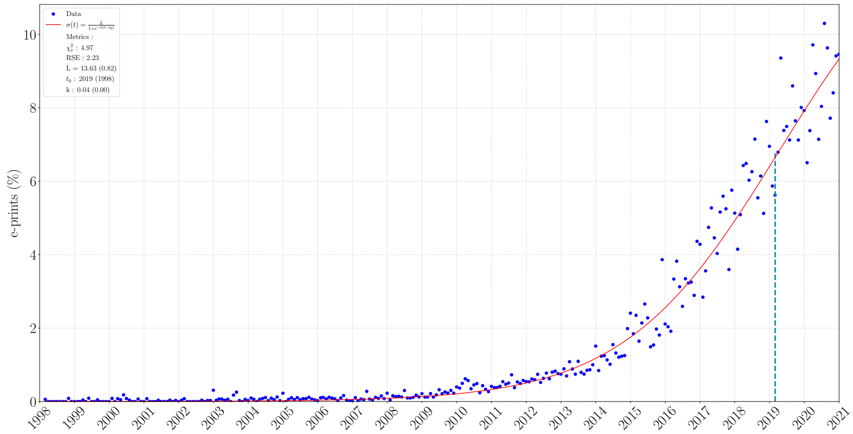

Figure 7: **Beginning of sigmoid growth of e-prints: computer vision and pattern recognition**. This figure depicts the normalized e-prints (blue) and the sigmoid fit (equation 2) is in red. We report the parameters of the sigmoid fit, the parameters, their standard errors in parenthesis and the χ_ν^2. The frequency is monthly. In contrast to Figure 5, the inflection point (*i.e*, t_0) has been reached around 2019.

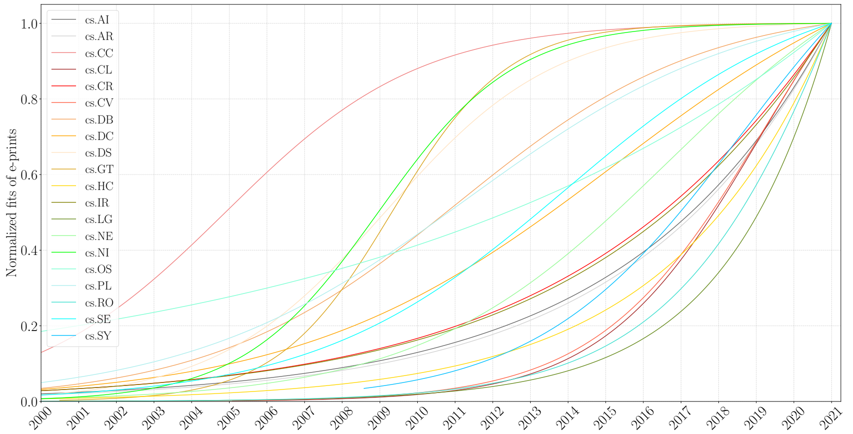

Figure 8: **Normalized fits of sigmoid functions**. The sigmoid fit is normalized through a division by its maximum. We plot all CSTCCs and the frequency is monthly. If we draw a 45° line starting from the upper-right corner (where all fits converge) to the bottom right, we can visually segregate CSTCCs that have reached their inflection point (above the line) from CTSCCs which have not reach their inflection point (under the line). Such a line is not represented on this plot for a better visualisation. Such a segregation can also be determined by analyzing t_0 of Table 3.

5.2. Security development is uncorrelated to technological development

We report the results of the tests of **H1b** in Table 4. With a multivariate time-series regression approach, we test the relation between the security development, $S_{x,t}$, and the technological development, $D_{x,t}$, across CSTCCs, x. There is no autocorrelation and heteroskedasticity of the error term u, and this for all x (hese tests are available upon request). We document a lack of statistically significant relation between $S_{x,t}$ and $D_{x,t}$: the statistical significance remains well above the 5% threshold for the great majority of regressors of D (*i.e.*, $q \ **= 0, \forall q, x$). Moreover, for the case of statistically significant regressors of D, their estimators are systematically of low magnitude and are well within the SE of regression. This makes us confident that technological development does not explain security development. Hence, **H1b** is verified for all CSTCCs.

		$S_{x,t}$				$D_{x,t}$			Adjusted R^2	SE regression	AIC	Sum resid^2	F-stat
		p_x	p_x^{**}	I_{S_x}	interpolated ratio	q_x	q_x^{**}	I_{D_x}					
$S_{x,t}$	`cs.AI`	12	7	1	0.04	1	0	2	0.419	0.099	-556.9	2.98	18.63
	`cs.AR`	12	8	1	0.23	2	2	1	0.448	0.139	-272.0	4.69	15.88
	`cs.CC`	12	10	1	0.24	1	0	1	0.478	0.060	-976.5	1.24	25.96
	`cs.CL`	12	8	1	0.02	1	0	2	0.436	0.094	-571.2	2.65	19.46
	`cs.CR`	12	5	1	0.12	1	0	1	0.508	0.095	-524.7	2.47	23.82
	`cs.CV`	12	8	1	0.15	1	0	3	0.569	0.111	-401.6	3.07	27.79
	`cs.DB`	12	7	1	0.07	1	1	1	0.480	0.113	-378.8	3.12	19.26
	`cs.DC`	12	8	1	0.03	1	0	1	0.508	0.104	-421.9	2.64	21.41
	`cs.DS`	12	8	1	0.22	1	0	1	0.419	0.075	-783.8	1.81	19.78
	`cs.GT`	12	8	1	0.11	1	0	1	0.403	0.107	-374.6	2.54	13.28
	`cs.HC`	12	10	2	0.14	1	0	2	0.739	0.140	-266.8	4.78	56.87
	`cs.IR`	11	6	1	0.06	1	0	2	0.624	0.115	-373.7	3.24	36.74
	`cs.LG`	12	8	1	0.09	1	0	2	0.420	0.105	-448.3	2.88	16.32
	`cs.NE`	12	7	1	0.11	2	0	1	0.549	0.111	-402.1	3.07	24.02
	`cs.NI`	12	9	1	0.09	7	4	1	0.398	0.111	-411.3	3.15	10.55
	`cs.OS`	12	4	1	0.35	2	1	1	0.431	0.147	-241.4	5.28	14.93
	`cs.PL`	12	8	1	0.20	2	0	1	0.437	0.099	-544.8	2.95	18.39
	`cs.RO`	12	5	1	0.23	1	0	1	0.350	0.104	-420.8	2.60	11.54
	`cs.SE`	12	8	1	0.11	2	0	1	0.475	0.105	-418.6	2.65	17.60
	`cs.SY`	12	4	1	0.26	1	0	2	0.379	0.079	-420.1	1.11	10.02

Table 4: **Multivariate time-series regression of security development**. This table lists the respective (i) autoregression order, p_x, and regression order, q_x, (ii) number of statistically significant (up to $p > 0.05$) autoregressors, p_x^{**}, and regressors, q_x^{**}, and (iii) the degree of differentiation, I_{S_x} and I_{D_x}. We also report the interpolated ratio of S, as these time-series cannot present null values (otherwise, the share of *security considerations* would be zero). To fulfill missing values, we perform a linear interpolation between concerned data points. Regression metrics are on the right.

5.3. *Security-attention* is increasing over time

Figure 9 shows the *rolling mean* (equ. 4) of the prevalence of *security considerations* (*i.e.*, the *security attention*).

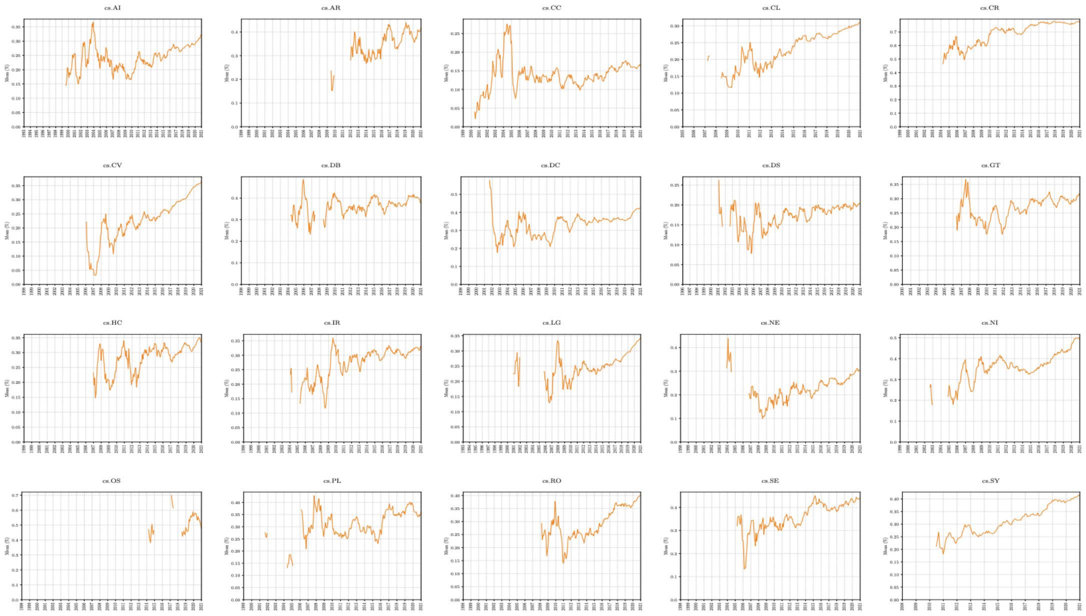

Figure 9: **Multi-plot of *security attention***

All CSTCCs display positive trends, depicting an increase in the prevalence of *security considerations* over time. Hence, **H2** is verified. We find empirical evidence of a *security-attention* pattern: the prevalence of *security considerations* grows for all CSTCCs over time – supporting the hypothesis that *security considerations* are gaining more attention at a later stage of technological development.

5.4. *Opinion* determined by *security considerations* and consensus

We report the results of the tests of **H3a** and **H3b** in Table 5. We test the relation between the instantaneous and aggregate measure of *opinion* and the prevalence of *security considerations* across CSTCCs (**H3a**) with a *Fama-Macbeth* approach. In all specifications, we document a significant relation that holds after the *Newey-West* adjustment (bandwidth ranging between 1 and 5, and truncated for the lag selection). More specifically, in the parsimonious version of the model, the prevalence of *security considerations* is significant at the 1% level (t-stats of 4.00 and 3.37 for the mean and the median, respectively). These results are robust to including the (log) number of e-prints and e-prints containing *security considerations*. The economic significance remains close and the statistical significance remains well below the 1% threshold. Thus, we rule out the possibility that our results are driven by either the numerator or the denominator used to construct the variable of interest. Moreover, given that we employ a cross-sectional methodology, these specifications also discard the possibility of a spurious time-effect as an explanation for our results. Interestingly, the point estimates for the two control variables are negative and significant at the 1% level in all but one specification (significant at the 5% level). This makes us confident that the prevalence of *security considerations* is different from the individual and absolute number of e-prints, either concerning the number of e-prints containing *security considerations* or the total number of e-prints. Hence, (**H3a**) is verified. The experts' *opinion* on a given CSTCC is positively related to their *security attention* expressed on the same CSTCC.

	Mean opinion								
Prevalence of *security considerations* ×10^4	2.54	2.69	3.98				1.36	1.22	-6.80
	(4.38)	(4.30)	(6.21)				(2.26)	(1.95)	(-4.64)
	[4.00]	[4.00]	[5.44]				[2.20]	[1.96]	[-2.73]
Opinion σ				-0.30	-0.22	-0.23	-0.25	-0.21	-0.16
				(-9.10)	(-7.17)	(-7.60)	(-7.89)	(-6.20)	(-4.31)
				[-8.82]	[-6.58]	[-7.00]	[-7.57]	[-5.81]	[-3.72]
Log (# e-prints with *security considerations*) ×10^4		-8.27			-6.40			-8.96	
		(-4.56)			(-4.02)			(-5.18)	
		[-5.14]			[-4.04]			[-5.03]	
Total			-22.00			-15.91			-41.50
			(-3.68)			(-3.69)			(-6.62)
			[-3.22]			[-3.27]			[-3.98]
Average R^2	0.05	0.12	0.11	0.13	0.19	0.19	0.19	0.25	0.34
	Median opinion								
Prevalence of *security considerations* ×10^4	1.90	1.20	2.99				1.11	0.92	-7.32.
	(3.54)	(3.39)	(4.72)				(1.98)	(1.56)	(-4.92)
	[3.37]	[3.21]	[4.24]				[1.95]	[1.61]	-[2.85]
Opinion σ				-0.14	-0.11	-0.12	-0.13	-0.10	-0.05
				(-5.15)	(-3.67)	(-4.05)	(-4.50)	(-3.18)	(-1.37)
				[-4.91]	[-3.41]	[-3.71]	[-4.24]	[-2.93]	[-1.18]
Log (# e-prints with *security considerations*) ×10^4		-6.44			-5.37			-8.04	
		(-3.86)			(-3.24)			(-4.46)	
		[-4.39]			[-2.86]			[-3.84]	
Total			-16.04			-12.13			-41.01
			(-2.80)			(-2.13)			(-6.55)
			[-2.27]			[-2.01]			[-3.78]
Average R^2	0.05	0.12	0.11	0.09	0.16	0.16	0.15	0.22	0.32

Table 5: **Cross-sectional regressions average of mean and median *opinion***. This Table reports the time-series average of parameters from cross-sectional regressions of the mean and the median *opinion*. The explanatory variables are the prevalence of *security considerations*, standard deviation of *opinion*, (log) number of e-prints with *security considerations*, and (log) number of e-prints. We report un-adjusted t-statistics in parenthesis, and Newey-West (1994) t-statistics in square brackets. We additionally report the average R^2, for each specification. The sample period is November 2002 – November 2020 and the number of time-series observations is 217.

In a final specification, we include the standard deviation of the sentiment (*i.e., opinion*) as an explanatory variable. We obtain similar orders of magnitude for the estimate and a statistical significance that remains at, or close to, the usual significance levels. When we consider this variable as a control, it does not modify the results. When we consider it as a variable of interest, the estimates are highly significant and negative (adjusted t-statistic up to −8.82 when the mean is used as a dependent variable). These results align with those of the financial literature. Such a negative relationship between *opinion* dispersion (of *e.g.*, analysts who provide price targets and recommendations for stocks) and actual stock returns is well documented [30] and explained by theoretical asset pricing models [48]. Hence, (**H3b**) is verified: The experts' *opinion* on a given CSTCC is positively related to the level of consensus on the same CSTCC. Consequently, we find empirical evidence for two contemporaneous determinants of the *opinion* towards the CSTCCs. First, the *opinion* expressed in e-prints is positively related to the prevalence of *security considerations*. Second, we find that consensus amongst the community is positively related to the *opinion* (standard deviation of the *opinion* is negatively related).

6. Discussion

Our results provide relevant insights in terms of practical implications and future work. Sub-section 6.1 offers practical implications for organizations undertaking acquisitions and investments in different *Computer-Science Technology Categories related to Cybersecurity* (CSTCCs). Sub-section 6.2 lists a promising research agenda for future work, notably for *security-dynamics* forecasting.

6.1. Practical implications

First, the sigmoid pattern of technological development – systematically observed for the great majority of the 20 CSTCCs under scrutiny – provides interesting hints for (i) benchmarking CSTCCs' development among each other, and (ii) subsequently forecasting the development of technologies in the short and medium terms. Knowing when a CSTCC is likely to reach its inflection point of growth is interesting to anticipate a CSTCC's development curve's slowing down. Subsequent analyses of such a decrease in growth may be used as an indicator of (i) technology maturity [80], and (ii) precaution to foresee potential technological obsolescence [71]. Such indicator may help to prioritize investment and acquisition of technologies with respect to their maturity and obsolescence levels. Making investment and acquisition choices related to technologies is always an opportunity-cost challenge. Knowing how a technology goes through different growth phases can provide considerable assistance in timing investment decisions appropriately [62]. Information about when and at which pace a CSTCC's growth is likely to slow down – also in comparison to other technologies developing concomitantly – is critical to set budget and investment and acquisition priorities. For instance, organizations willing to invest in emerging technologies are likely to prioritize technologies that have not reached their inflection point, will organizations willing to acquire technologies are likely to prioritize mature technologies that are relatively far from reaching their obsolescence.

Second, the lack of a significant relationship between technological development and security development – as well as the fact that *security considerations* arrive at a later stage of technological development – confirms previous intuitions about how ill-defined security development is among computer-science technologies [19, 51, 16, 6, 70]. Such empirical findings put to the fore indicators of how important security is considered among the technological development pipeline. More practically speaking, such findings are worth considering for benchmarking purposes to evaluate where the security development stands among different technologies – a contemporary question that is gaining momentum within the technology domain [7]. Decision-makers who invest or acquire technologies must grasp how security is considered within technologies of interest. This aspect is critical, especially when considering that security has often been sacrificed for revenue or user-base growth [7]. How much attention has been paid to *security considerations* during technological development? When did these *security considerations* start getting momentum? Our approach recognizes that such security-related aspects are uncertain without proper measurements. Therefore, we modeled these aspects to shed light on *security considerations*. Based on theses aspects, our security indicators may help IT decision-makers (such as CISOs) to prioritize investment or acquisition. Under a societal change perspective, such findings give some relevant thinking grounds to how aspects related to security are effectively envisioned and considered within the overall engineering process. The question of how important security development is among economic and technological development factors – and how such a question might be considered going forward – constitutes a conceptual debate that is worth reconsidering.

Third, to complement the investigation of the technological and security developments *per se*, we also assessed experts' *opinion* on technology. Our approach enables us to investigate the extent to which *security considerations* affect (i) the *opinion*, and (ii) the *opinion* consensus towards a given CSTCC. Again, such metrics can then be used as a benchmark among CSTCCs – an indicator relevant for decision-making related to acquisitions and investments in specific technologies. For instance, similarly to financial assets, the *opinion* given to specific technologies may reveal crucial in *pricing* these same technologies based on the perceived security risks. There is little doubt that investment and acquisition insights and subsequent technology pricing models can be derived from such results. The relation of the standard deviation of *opinion* with that of the *security considerations* is another striking example that security risk associated with technology development can be priced thanks to *opinion* mining. The consistency and the stability of the reported metrics over short to medium periods are essential for technology investment and acquisition.

6.2. Limitations and future work

We measured the amount of scientific work produced through time as a proxy for technology development according to abundant bibliometric literature (*e.g.*, [31, 45, 77]). More specifically, we measured the number of e-prints uploaded through time in the `arXiv` repository for each CSTCC. With this measure, we aimed to capture the amount of attention that a given community (the scientific community) provides to a given CSTCC. In the field of computer science, it is common practice to upload e-prints on the `arXiv` repository whenever such e-prints are ready for submission in a scientific conference (or sometimes in scientific journals). The `arXiv` repository is thus considered by computer scientists as a central repository for academic research related to their field. Consequently, we argue that the most scientific advances and their development are captured through the `arXiv` platform. However, other providers such as `Microsoft Academic`, `Scopus`, `Web of Science`, `Semantic Scholar`, or the promising API of `OpenAlex` might be considered complementary platforms to capture an even more complete picture of scientific works related to computer science.[17] However, future work aiming to integrate such data may face non-trivial information retrieval and data integration challenges.

Also, other aspects and measures of technological development may be used, such as technology adoption (*e.g.*, by considering the growth in software/hardware instances and the number of users, and/or by considering the number of patents and the dynamics of social-media heuristic recurrences), or technology maturity (*e.g.*, by considering the *opinion* of users in their reviews, and/or TRL measurements). In this work, we focused on the scientific community's attention to different CSTCCs through the analysis of a subset of scholarly literature. Future work might include such indicators and compare their trends with ours to assess potential idiosyncrasies between indicators.

Concerning the prevalence of *security considerations* expressed in e-prints, we selected a set of keywords related to the *key concepts* depicted in the *Information Security* portal of *Wikipedia*. These key concepts relate to the well-known CIA triad (*i.e.*, *confidentiality*, *integrity*, and *availability*) and the *non-repudiation* principle [23, 78]. Our results support the hypothesis that *security attention* is gaining momentum at a later stage of technological development. We leave for future research the investigation and measure of how substantial this delay between technological development and security development is. This could be done by implementing a delay function, which could also be employed to investigate whether a *catch-up effect* is present and for which CSTCCs. However, one might argue that such a pattern in *security attention* is induced by an omitted variable, the general hype in cybersecurity which grows over time. Unfortunately, accounting for such an omitted variable seems hardly feasible in practice. Yet, when we conducted the Fama-MacBeth (cross-sectional analysis), which is not influenced by any time trend – and thus by such an omitted variable –, we found support for the relation between *security attention* and *opinion*. This last fact demonstrates that such a potential omitted variable does not affect **H3a**. Future research could enhance the selection of keywords related to *security considerations* by implementing other information retrieval methods such as *tf-idf* or *Key-BERT* to capture the recurrences of words related to security, and then use the most recurrent ones as filters for capturing *security considerations* in e-prints.

Additionally, our classical lexicon-based approach to capture *opinion* – applied through a standard cumulative-sentiment function – may be enhanced by applying more sophisticated machine-learning approaches (*e.g.*, supervised decision-tree classifiers, linear classifiers using support-vector machines or neural networks, rule-based classifiers, or probabilistic classifiers involving naive Bayes or maximum entropy principles). Such approaches would yield a higher precision for sentiment analysis through semantics and heuristics. However, researchers willing to use such more sophisticated NLP methods may face issues finding labeled datasets. For instance, *BERT* [87] and *XLNet* [66] models, despite being pre-trained with a plethora of datasets, are not directly transferable to datasets presenting other text structures. Unfortunately, to the best of our knowledge, there are no academic-work datasets labeled for sentiment analysis.

[17] `https://docs.openalex.org/api`

7. Conclusion

Little work has been done to model a holistic and dynamic indicator that captures the overall security development of technologies – especially with respect to technological development. We conceptualized and measured such an indicator based on the investigation of what we call *security dynamics*, constituted by (i) the statistical relation between technological and security developments, (ii) the security development, modeled as the evolution of *security considerations* among technologies, and (iii) the effect of security development on the *opinion* given to technologies by experts who produce and evaluate these same technologies. We adopted a bibliometric approach related to 20 computer-science technology categories.

We found results that together bring a unique perspective on the critical question how security evolves as part of technological development. First, there is a lack of relationship between technological and security developments. Second, security is gaining more attention at a later stage of technological development. Third, the experts' *opinion* related to each technology is explained by the prevalence of *security considerations*.

These results bring new methods for understanding, modeling, and benchmarking *security dynamics* of computer-science technologies. In turn, these methods and models open new heuristics for considering changes related to the security of information systems.

References

[1] Abernathy, W.J., Utterback, J.M., et al., 1978. Patterns of industrial innovation. Technology review 80, 40–47.

[2] Adamuthe, A.C., Tomke, J.V., Thampi, G.T., 2014. Technology Forecasting: The Case of Cloud Computing and Sub-Technologies. International Journal of Computer Applications 106. Publisher: Citeseer.

[3] Adner, R., Levinthal, D.A., 2002. The emergence of emerging technologies. California management review 45, 50–66.

[4] Akerlof, G.A., 1978. The market for "lemons": Quality uncertainty and the market mechanism, in: Uncertainty in economics. Elsevier, pp. 235–251.

[5] Andersen, B., 1999. The hunt for S-shaped growth paths in technological innovation: a patent study*. Journal of Evolutionary Economics 9, 487–526. doi:10.1007/s001910050093.

[6] Anderson, R., 2001. Why information security is hard-an economic perspective, in: Seventeenth Annual Computer Security Applications Conference, IEEE. pp. 358–365.

[7] Anderson, R., 2020. Security Engineering: A Guide to Building Dependable Distributed Systems. 3rd edition ed., Wiley.

[8] Anderson, R., Moore, T., 2006. The economics of information security. science 314, 610–613. Publisher: American Association for the Advancement of Science.

[9] Anderson, R., Moore, T., 2007. The economics of information security: A survey and open questions, in: Fourth bi-annual Conference on the Economics of the Software and Internet Industries, pp. 19–20.

[10] Assante, M.J., Tobey, D.H., 2011. Enhancing the cybersecurity workforce. IT professional 13, 12–15. Publisher: IEEE.

[11] Asteriou, D., Hall, S.G., 2015. Applied econometrics. Macmillan International Higher Education.

[12] Bai, X., 2011. Predicting consumer sentiments from online text. Decision Support Systems 50, 732–742. doi:10.1016/j.dss.2010.08.024.

[13] Bengisu, M., Nekhili, R., 2006. Forecasting emerging technologies with the aid of science and technology databases. Technological Forecasting and Social Change 73, 835–844. doi:10.1016/j.techfore.2005.09.001.

[14] Bevington, P.R., Robinson, D.K., 2003. Data reduction and error analysis. McGraw Hill, New York .

[15] Brock, G.W., 2021. The second information revolution. Harvard University Press.

[16] Böhme, R. (Ed.), 2013. The Economics of Information Security and Privacy. Springer-Verlag, Berlin Heidelberg. doi:10.1007/978-3-642-39498-0.

[17] Calleja-Sanz, G., Olivella-Nadal, J., Solé-Parellada, F., 2020. Technology Forecasting: Recent Trends and New Methods. Research Methodology in Management and Industrial Engineering , 45–69Publisher: Springer.

[18] Carlton, M., 2016. Development of a cybersecurity skills index: A scenarios-based, hands-on measure of non-IT professionals' cybersecurity skills. PhD Thesis. Nova Southeastern University.

[19] Casola, V., De Benedictis, A., Rak, M., Villano, U., 2020. A novel Security-by-Design methodology: Modeling and assessing security by SLAs with a quantitative approach. Journal of Systems and Software 163, 110537. Publisher: Elsevier.

[20] Chang, W.L., Wang, J.Y., 2018. Mine is yours? Using sentiment analysis to explore the degree of risk in the sharing economy. Electronic Commerce Research and Applications 28, 141–158. doi:10.1016/j.elerap.2018.01.014.

[21] Chen, H., Zhang, G., Zhu, D., Lu, J., 2017. Topic-based technological forecasting based on patent data: A case study of Australian patents from 2000 to 2014. Technological Forecasting and Social Change 119, 39–52. doi:10.1016/j.techfore.2017.03.009.

[22] Chen, Y.H., Chen, C.Y., Lee, S.C., 2011. Technology forecasting and patent strategy of hydrogen energy and fuel cell technologies. International Journal of Hydrogen Energy 36, 6957–6969. doi:10.1016/j.ijhydene.2011.03.063.

[23] Cherdantseva, Y., Hilton, J., 2013. A Reference Model of Information Assurance Security, in: 2013 International Conference on Availability, Reliability and Security, pp. 546–555. doi:10.1109/ARES.2013.72.

[24] Choi, J., Hwang, Y.S., 2014. Patent keyword network analysis for improving technology development efficiency. Technological Forecasting and Social Change 83, 170–182. doi:10.1016/j.techfore.2013.07.004.

[25] Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I., 1990. Stl: A seasonal-trend decomposition. J. Off. Stat 6, 3–73.

[26] Coccia, M., 2005. Technometrics: Origins, historical evolution and new directions. Technological Forecasting and Social Change 72, 944–979. doi:10.1016/j.techfore.2005.05.011.

[27] Daim, T., Iskin, I., Li, X., Zielsdorff, C., Bayraktaroglu, A.E., Dereli, T., Durmusoglu, A., 2012. Patent analysis of wind energy technology using the patent alert system. World Patent Information 34, 37–47. doi:10.1016/j.wpi.2011.11.001.

[28] Daim, T., Yalçin, H., 2022. Digital Transformations: New Tools and Methods for Mining Technological Intelligence. Edward Elgar Publishing, Incorporated. URL: https://books.google.ch/books?id=po-MzgEACAAJ.

[29] Daim, T.U., Chiavetta, D., Porter, A.L., Saritas, O., 2016. Anticipating Future Innovation Pathways Through Large Data Analysis. Springer.

[30] Diether, K.B., Malloy, C.J., Scherbina, A., 2002. Differences of Opinion and the Cross Section of Stock Returns. The Journal of Finance 57, 2113–2141. Publisher: [American Finance Association, Wiley].

[31] Dotsika, F., Watkins, A., 2017. Identifying potentially disruptive trends by means of keyword network analysis. Technological Forecasting and Social Change 119, 114–127. Publisher: Elsevier.

[32] Dou, R., Zhang, Y., Nan, G., 2017. Iterative product design through group opinion evolution. International Journal of Production Research 55, 3886–3905. doi:10.1080/00207543.2017.1316020.

[33] Fama, E.F., MacBeth, J.D., 1973. Risk, Return, and Equilibrium: Empirical Tests. Journal of Political Economy 81, 607–636. Publisher: University of Chicago Press.

[34] Fang, X., Zhan, J., 2015. Sentiment analysis using product review data. Journal of Big Data 2, 5. doi:10.1186/s40537-015-0015-2.

[35] Golembiewski, B., vom Stein, N., Sick, N., Wiemhöfer, H.D., 2015. Identifying trends in battery technologies with regard to electric mobility: evidence from patenting activities along and across the battery value chain. Journal of Cleaner Production 87, 800–810. doi:10.1016/j.jclepro.2014.10.034.

[36] Goode, J., Levy, Y., Hovav, A., Smith, J., 2018. Expert assessment of organizational cybersecurity programs and development of vignettes to measure cybersecurity countermeasures awareness. Online Journal of Applied Knowledge Management (OJAKM) 6, 54–66.

[37] Guo, W., Wang, H., Tian, Y., Xian, M., 2019. Research on" Cyberspace Security Testing and Evaluation" Technology Development Trend, in: 2019 International Conference on Communications, Information System and Computer Engineering (CISCE), IEEE. pp. 363–367.

[38] Gurung, A., Raja, M., 2016. Online privacy and security concerns of consumers. Information & Computer Security 24, 348–371. doi:10.1108/ICS-05-2015-0020. publisher: Emerald Group Publishing Limited.

[39] Haleem, A., Mannan, B., Luthra, S., Kumar, S., Khurana, S., 2019. Technology forecasting (TF) and technology assessment (TA) methodologies: a conceptual review. Benchmarking: An International Journal 26, 48–72. doi:10.1108/BIJ-04-2018-0090. publisher: Emerald Publishing Limited.

[40] Han, J., Moraga, C., 1995. The influence of the sigmoid function parameters on the speed of backpropagation learning, in: Mira, J., Sandoval, F. (Eds.), From Natural to Artificial Neural Computation, Springer, Berlin, Heidelberg. pp. 195–201. doi:10.1007/3-540-59497-3_175.

[41] Hao, J., Yan, Y., Gong, L., Wang, G., Lin, J., 2014. Knowledge map-based method for domain knowledge browsing. Decision Support Systems 61, 106–114. doi:10.1016/j.dss.2014.02.001.

[42] Howard, M., Lipner, S., 2006. The security development lifecycle. volume 8. Microsoft Press Redmond.

[43] Huang, J., Boh, W.F., Goh, K.H., 2019. Opinion convergence versus polarization: examining opinion distributions in online word-of-mouth. Journal of the Association for Information Science and Technology 70, 1183–1193. doi:https://doi.org/10.1002/asi.24193. _eprint: https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.24193.

[44] Hubbard, D.W., Seiersen, R., 2016. How to measure anything in cybersecurity risk. John Wiley & Sons.

[45] Jaewoo, C., Woonsun, K., 2014. Themes and Trends in Korean Educational Technology Research: A Social Network Analysis of Keywords. Procedia - Social and Behavioral Sciences 131, 171–176. doi:10.1016/j.sbspro.2014.04.099.

[46] Jaffe, A.B., Newell, R.G., Stavins, R.N., 2002. Environmental policy and technological change. Environmental and resource economics 22, 41–70.

[47] Jang-Jaccard, J., Nepal, S., 2014. A survey of emerging threats in cybersecurity. Journal of Computer and System Sciences 80, 973–993. Publisher: Elsevier.

[48] Johnson, T.C., 2004. Forecast Dispersion and the Cross Section of Expected Returns. The Journal of Finance 59, 1957–1978. doi:https://doi.org/10.1111/j.1540-6261.2004.00688.x.

[49] Jun, S., Sung Park, S., Sik Jang, D., 2012. Technology forecasting using matrix map and patent clustering. Industrial Management & Data Systems 112, 786–807. doi:10.1108/02635571211232352. publisher: Emerald Group Publishing Limited.

[50] Klepper, S., 1997. Industry life cycles. Industrial and corporate change 6, 145–182.

[51] Kreitz, M., 2019. Security by design in software engineering. ACM SIGSOFT Software Engineering Notes 44, 23–23. Publisher: ACM New York, NY, USA.

[52] Laube, S., Böhme, R., 2017. Strategic aspects of cyber risk information sharing. ACM Computing Surveys (CSUR) 50, 1–36. Publisher: ACM New York, NY, USA.

[53] Lee, C., 2021. A review of data analytics in technological forecasting. Technological Forecasting and Social Change 166, 120646. doi:10.1016/j.techfore.2021.120646.

[54] Lee, P.C., Su, H.N., Wu, F.S., 2010. Quantitative mapping of patented technology — The case of electrical conducting polymer nanocomposite. Technological Forecasting and Social Change 77, 466–478. doi:10.1016/j.techfore.2009.08.006.

[55] Lehrer, K., Wagner, C., 2012. Rational consensus in science and society: A philosophical and mathematical study. volume 24. Springer Science & Business Media.

[56] Li, L., He, W., Xu, L., Ash, I., Anwar, M., Yuan, X., 2019a. Investigating the impact of cybersecurity policy awareness on employees' cybersecurity behavior. International Journal of Information Management 45, 13–24. Publisher: Elsevier.

[57] Li, X., Xie, Q., Daim, T., Huang, L., 2019b. Forecasting technology trends using text mining of the gaps between science and technology: The case of perovskite solar cell technology. Technological Forecasting and Social Change 146, 432–449. doi:10.1016/j.techfore.2019.01.012.

[58] Liu, B., 2012. Sentiment Analysis and Opinion Mining.

[59] Liu, Y., Sarabi, A., Zhang, J., Naghizadeh, P., Karir, M., Bailey, M., Liu, M., 2015. Cloudy with a chance of breach: Forecasting cyber security incidents, in: 24th ${$USENIX$}$ Security Symposium (${$USENIX$}$ Security 15), pp. 1009–1024.

[60] Lotfi, A., Lotfi, A., Halal, W.E., 2014. Forecasting technology diffusion: a new generalisation of the logistic model. Technology Analysis & Strategic Management 26, 943–957. doi:10.1080/09537325.2014.925105. publisher: Routledge _eprint: https://doi.org/10.1080/09537325.2014.925105.

[61] Maks, I., Vossen, P., 2013. Sentiment Analysis of Reviews: Should we analyze writer intentions or reader perceptions?, in: Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013, INCOMA Ltd. Shoumen, BULGARIA, Hissar, Bulgaria. pp. 415–419.

[62] Marcus Matthias Keupp, Percia David, D., Mermoud, Alain, 2019. Militärökonomie. Springer.

[63] Meland, P.H., Tokas, S., Erdogan, G., Bernsmed, K., Omerovic, A., 2021. A systematic mapping study on cyber security indicator data. Electronics 10, 1092.

[64] Mikheev, A.V., 2020. Technological forecasting related to the energy sector: a scientometric overview. E3S Web of Conferences 209, 02022. doi:10.1051/e3sconf/202020902022.

[65] Moré, J.J., 1978. The Levenberg-Marquardt algorithm: implementation and theory, in: Numerical analysis. Springer, pp. 105–116.

[66] Myagmar, B., Li, J., Kimura, S., 2019. Cross-domain sentiment classification with bidirectional contextualized transformer language models. IEEE Access 7, 163219–163230. Publisher: IEEE.

[67] Newey, W.K., West, K.D., 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 55, 703–708. doi:10.2307/1913610. publisher: [Wiley, Econometric Society].

[68] Newey, W.K., West, K.D., 1994. Automatic Lag Selection in Covariance Matrix Estimation. The Review of Economic Studies 61, 631–653. doi:10.2307/2297912.

[69] Noh, H., Song, Y.K., Lee, S., 2016. Identifying emerging core technologies for the future: Case study of patents published by leading telecommunication organizations. Telecommunications Policy 40, 956–970. doi:10.1016/j.telpol.2016.04.003.

[70] Panarello, A., Tapas, N., Merlino, G., Longo, F., Puliafito, A., 2018. Blockchain and iot integration: A systematic survey. Sensors 18, 2575. Publisher: Multidisciplinary Digital Publishing Institute.

[71] Parvin Jr, A.J., 2017. Forecasting technology obsolescence: Assessing the existing literature, a systematic review, in: Proceedings of the International Annual Conference of the American Society for Engineering Management., American Society for Engineering Management (ASEM). pp. 1–13.

[72] Perez, C., 2010. Technological revolutions and techno-economic paradigms. Cambridge journal of economics 34, 185–202.

[73] Perez, C., et al., 2010. The financial crisis and the future of innovation: A view of technical change with the aid of history. Technical Report. TUT Ragnar Nurkse Department of Innovation and Governance.

[74] Pletea, D., Vasilescu, B., Serebrenik, A., 2014. Security and emotion: sentiment analysis of security discussions on github, in: Proceedings of the 11th working conference on mining software repositories, pp. 348–351.

[75] Porter, A.L., Roper, A.T., Mason, T.W., Rossini, F.A., Banks, J., 2011. Forecasting and Management of Technology. John Wiley & Sons. Google-Books-ID: L4Bqo1HHq7UC.

[76] Priestley, M., Sluckin, T.J., Tiropanis, T., 2020. Innovation on the web: the end of the s-curve? Internet Histories 4, 390–412.

[77] Rezaeian, M., Montazeri, H., Loonen, R.C.G.M., 2017. Science foresight using life-cycle analysis, text mining and clustering: A case study on natural ventilation. Technological Forecasting and Social Change 118, 270–280. doi:10.1016/j.techfore.2017.02.027.

[78] Ritzdorf, H., Wüst, K., Gervais, A., Felley, G., Capkun, S., 2017. Tls-n: Non-repudiation over tls enabling-ubiquitous contentsigning for disintermediation, in: TLS-N: Non-repudiation over TLS Enabling-Ubiquitous ContentSigning for Disintermediation, ETH Zurich. pp. 1–16.

[79] Rogers, E.M., 1995. Diffusion of Innovations: modifications of a model for telecommunications, in: Die diffusion von innovationen in der telekommunikation. Springer, pp. 25–38.

[80] Rogers, E.M., 2010. Diffusion of innovations. Simon and Schuster.

[81] Santos, J.C.S., Tarrit, K., Mirakhorli, M., 2017. A Catalog of Security Architecture Weaknesses, in: 2017 IEEE International Conference on Software Architecture Workshops (ICSAW), pp. 220–223. doi:10.1109/ICSAW.2017.25.

[82] Serrano-Guerrero, J., Olivas, J.A., Romero, F.P., Herrera-Viedma, E., 2015. Sentiment analysis: A review and comparative analysis of web services. Information Sciences 311, 18–38.

[83] Shalf, J., 2020. The future of computing beyond Moore's law. Philosophical Transactions of the Royal Society A 378, 20190061. Publisher: The Royal Society Publishing.

[84] Son, H., Kim, C., Kim, H., Han, S.H., Kim, M.K., 2010. Trend analysis of research and development on automation and robotics technology in the construction industry. KSCE Journal of Civil Engineering 14, 131–139. doi:10.1007/s12205-010-0131-7.

[85] Steinmetz, N., 2011. Rational Iteration: Complex Analytic Dynamical Systems. Walter de Gruyter. Google-Books-ID: qZGWgVuGHiYC.

[86] Sutton, C., Gong, L., 2017. Popularity of arxiv. org within computer science. arXiv preprint arXiv:1710.05225 .

[87] Tenney, I., Das, D., Pavlick, E., 2019. BERT rediscovers the classical NLP pipeline. arXiv preprint arXiv:1905.05950 .

[88] Verhulst, P.F., 1838. Notice sur la loi que la population suit dans son accroissement. Corresp. Math. Phys. 10, 113–126.

[89] Yang, J., Sarathy, R., Lee, J., 2016. The effect of product review balance and volume on online Shoppers' risk perception and purchase intention. Decision Support Systems 89, 66–76. doi:10.1016/j.dss.2016.06.009.

[90] Yang, Y., Liu, Y., Li, H., Yu, B., 2015. Understanding perceived risks in mobile payment acceptance. Industrial Management & Data Systems 115, 253–269. doi:10.1108/IMDS-08-2014-0243. publisher: Emerald Group Publishing Limited.

[91] Yüzügüllü, E., Deason, J.P., 2007. Structuring objectives to facilitate convergence of divergent opinion in hydrogen production decisions. Energy Policy 35, 452–460. doi:10.1016/j.enpol.2005.12.001.

[92] Zharov, V.S., Kozlov, A.V., 2018. Management of technological development of enterprises on the basis of a life cycle model, in: Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS), IEEE. pp. 181–184.

Appendix

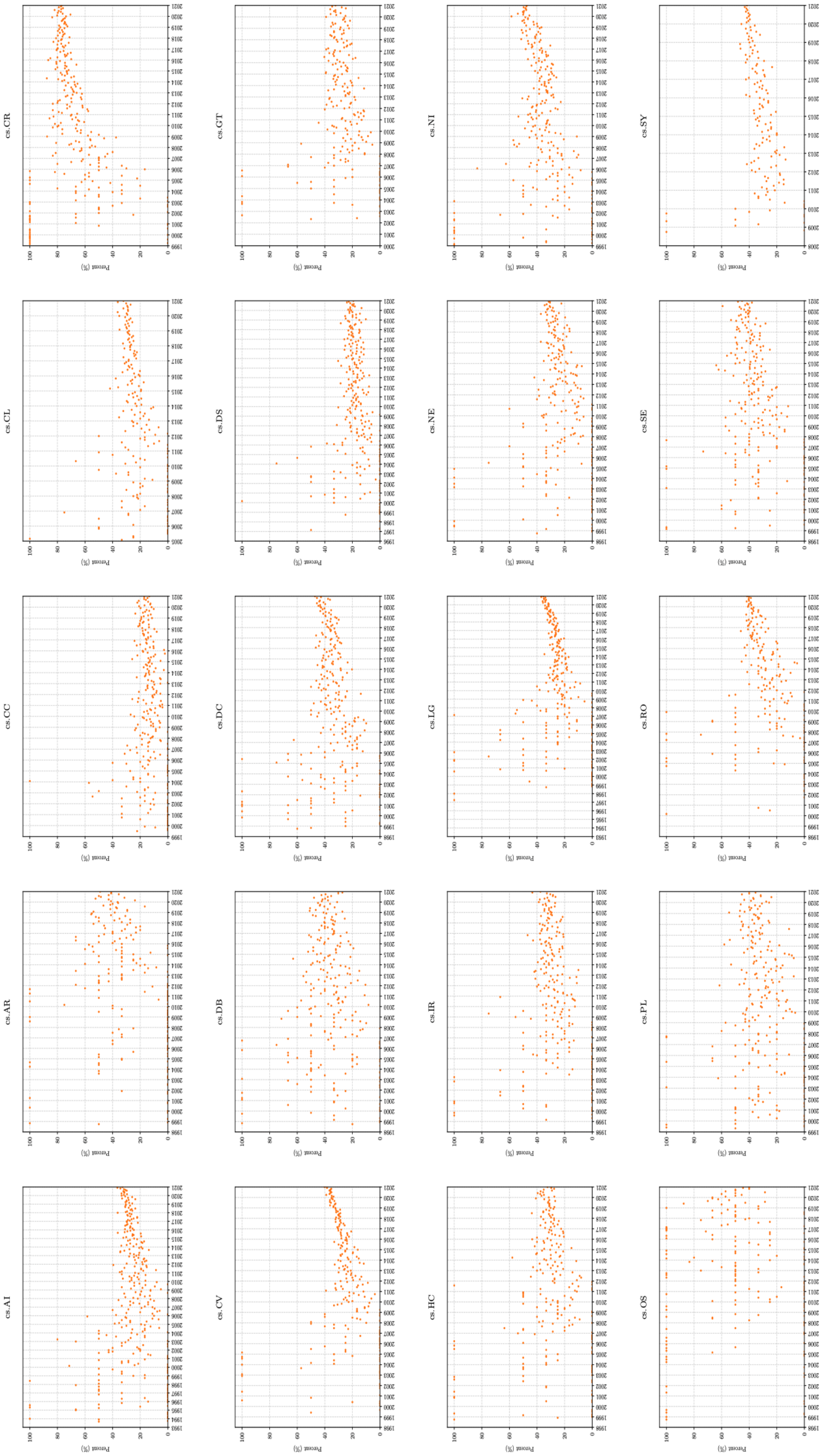

Figure 10: **Multi-plot of *security considerations***

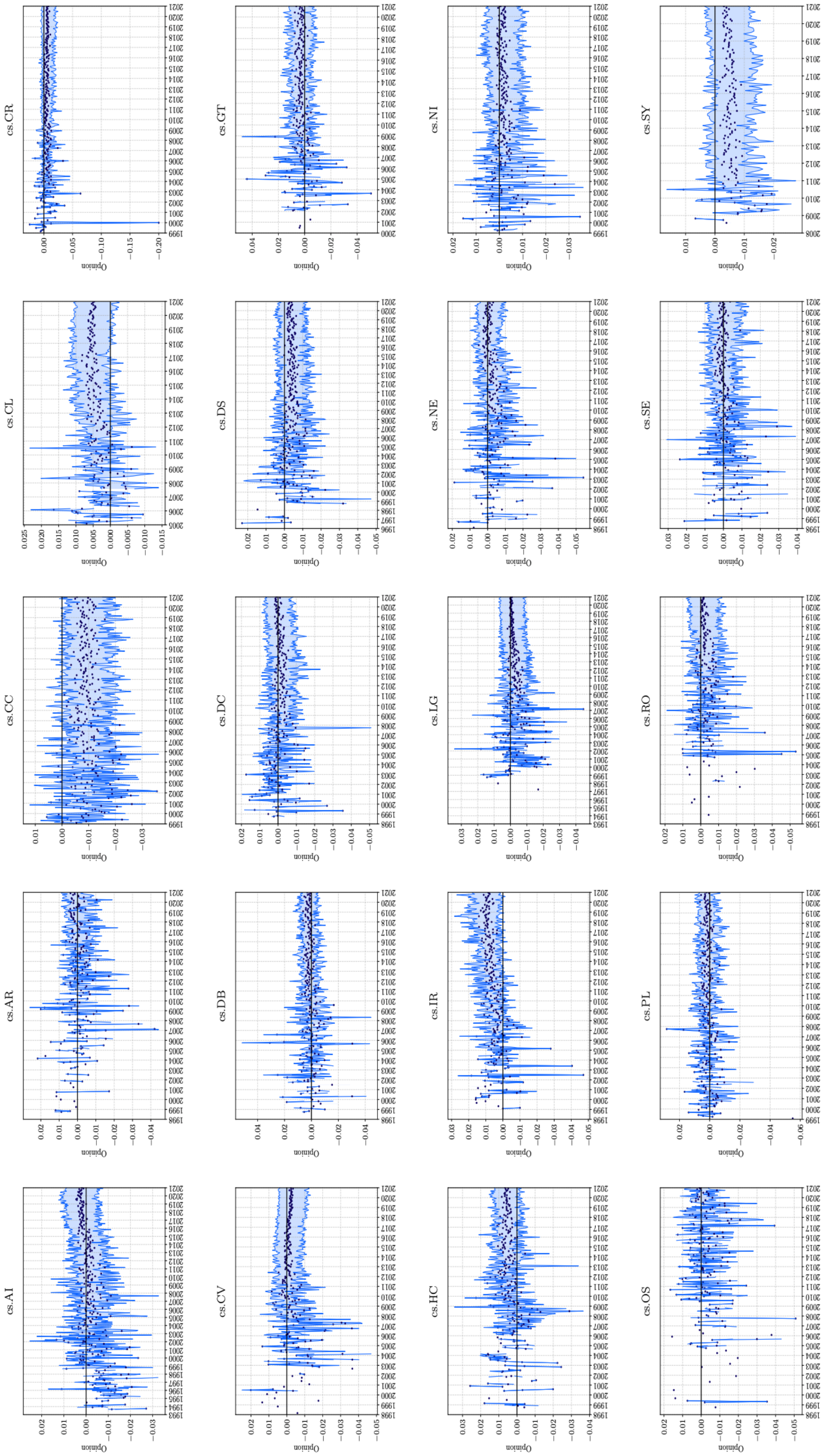

Figure 11: **Multi-plot of *opinion***

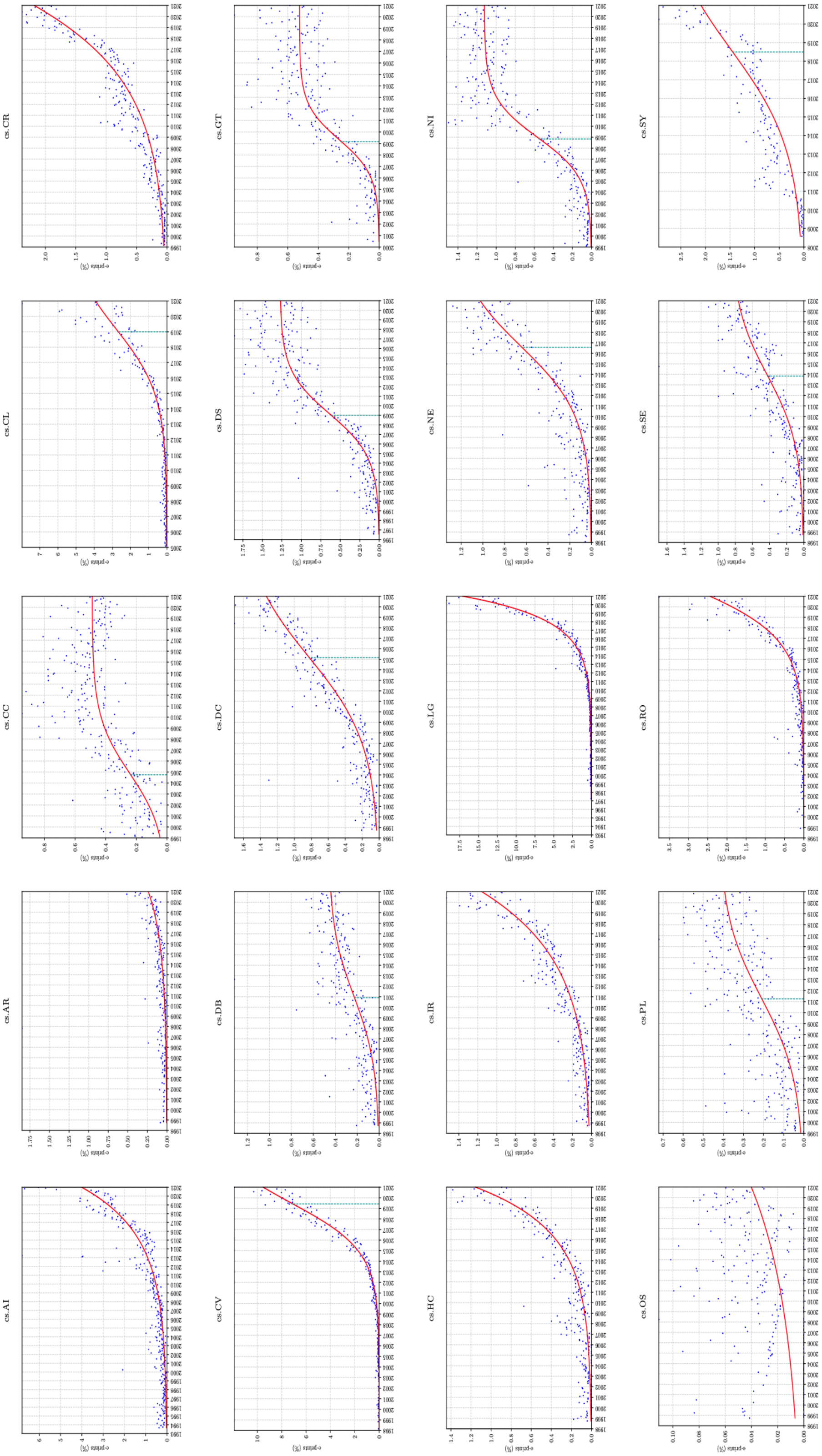

Figure 12: **Multi-plot of sigmoid fits**