S-curves Demystified: Empirical Evidence of Multi-Sigmoid Development in Computer-Science Technologies

Technological development is the invention, innovation, and diffusion of new technologies, i.e., novel processes that produce outputs that were not feasible before [1]. An invention constitutes a scientifically or technologically new product or process, which may or may not develop into an innovation, which happens when an invention is commercialized. A successful innovation becomes gradually more adopted by the market and society. This diffusion is a gradual process that can be modeled by S-curves, which have a long history and appear in many different domains, e.g., Biology [2]. S-curve segment development into three phases: introduction, growth, and maturation phases.

Multiple research works model technological development with S-curves with a single growth phase (single-sigmoid) ([4], [5], [6]). The carrying capacity refers to the maximum population size a biological system can support. We can apply this concept also to technological developments; here, the carrying capacity caps the amount of technological development possible. For single-sigmoids, the carrying capacity stays constant. However, the carrying capacity of a system developing technologies can often change, e.g., a new application of a technology is discovered, or a technology diffuses at different rates in different domains [7]. Thus, a technology can go through a second growth phase [8] or multiple growth phases [7]. In the following, we use the number of distinct increases in the carrying capacity to refer to technologies that go through one, two, or multiple growth spurts as single-, double-, or multi-sigmoid growth. Meyer models the number of nuclear explosions, the population of Japan, the cumulative number of U.S. universities, and the adoption of electric generators as bi- or multi-sigmoid growth [8], [7], Kucharavy and De Guio model energy consumption and infrastructure development as multi-sigmoid growth [6].

Percia et al. [3] show the prevalence of single-sigmoid growth patterns in computer-science subcategories on the e-print archive `arXiv`. However, we hypothesize that the development of some subcategories exhibits multiple bouts of growth that single-sigmoid growth cannot capture. As previously shown, past discoveries of multi-sigmoid patterns were in vast domains, e.g., infrastructure development, but not in smaller-scale computer-science technologies. Thus, the confirmation of multi-sigmoid growth patterns in computer science technologies remains to be done.

In this work, we use the e-prints uploaded to the `arXiv` repository as our dataset. `arXiv` is an open-

arXiv	subcategory name	# e-prints	χ_v single	χ_v multi	χ_v Difference Test p-value	RMSE single	RMSE multi
cs.AI	Artificial Intelligence	38 620	11.8	8.9	3.95×10^{-53}	6.59×10^{-3}	5.73×10^{-3}
cs.AR	Hardware Architecture	2573	1.9	1.6	4.15×10^{-4}	1.38×10^{-3}	1.36×10^{-3}
cs.CC	Computational Complexity	8492	2.3	2.4	1	1.36×10^{-3}	1.36×10^{-3}
cs.CL	Computation and Language	29 528	14.1	13.8	1.16×10^{-5}	6.4×10^{-3}	6.27×10^{-3}
cs.CR	Cryptography and Security	19 784	2.8	2	7.16×10^{-12}	1.81×10^{-3}	1.44×10^{-3}
cs.CV	Computer Vision and Pattern Recognition	64 696	6	5.8	9.81×10^{-5}	5.43×10^{-3}	5.38×10^{-3}
cs.DB	Databases	6269	2.4	2.4	1.99×10^{-1}	1.25×10^{-3}	1.23×10^{-3}
cs.DC	Distributed, Parallel, and Cluster Computing	14 955	2.2	2.1	4.82×10^{-2}	1.54×10^{-3}	1.49×10^{-3}
cs.DS	Data Structures and Algorithms	18 269	3.1	3.1	5.37×10^{-2}	2.12×10^{-3}	2.1×10^{-3}
cs.GT	Computer Science and Game Theory	7992	2.1	2.1	5.63×10^{-1}	1.13×10^{-3}	1.12×10^{-3}
cs.HC	Human-Computer Interaction	8774	2.2	2.2	1.45×10^{-1}	1.19×10^{-3}	1.16×10^{-3}
cs.IR	Information Retrieval	10 407	2.3	2.1	7.23×10^{-4}	1.24×10^{-3}	1.14×10^{-3}
cs.LG	Machine Learning	94 024	17.1	10.6	5.89×10^{-103}	1.16×10^{-2}	9.14×10^{-3}
cs.NE	Neural and Evolutionary Computing	10 155	3	2.9	7.57×10^{-3}	1.54×10^{-3}	1.49×10^{-3}
cs.NI	Networking and Internet Architecture	16 606	2.3	2.4	1	1.67×10^{-3}	1.67×10^{-3}
cs.OS	Operating Systems	652	1	0.9	7.85×10^{-1}	2.92×10^{-4}	2.98×10^{-4}
cs.PL	Programming Languages	5731	2.6	2.4	1.52×10^{-2}	1.11×10^{-3}	1.07×10^{-3}
cs.RO	Robotics	16 187	4.2	4.1	5.74×10^{-2}	2.69×10^{-3}	2.68×10^{-3}
cs.SE	Software Engineering	10 032	4.1	3.5	4.98×10^{-8}	1.94×10^{-3}	1.80×10^{-3}
cs.SY	Systems and Control	18 347	7.7	5.1	4.53×10^{-23}	3.45×10^{-3}	3.11×10^{-3}

Table 1: **20 selected computer-science subcategories on `arXiv` analyzed by Percia et al. [3]**. The authors confirmed that these categories follow a sigmoid-growth pattern, shown by χ_v single sigmoid values close to 1. Subcategories with $\chi_v \gg 1$ (`cs.AI`, `cs.LG`, `cs.CL`) did not reach their inflection point yet. Subcategories highlighted with gray are subcategories for which we determined visually and using the χ^2 difference test that the multi-sigmoid describes their development better. The difference is significant if p-value < 0.05. An exception are the subcategories `cs.AI`, `cs.LG`, and `cs.CL`, where we determined visually that the multi-sigmoid overfits. This leaves nine out of 20 subcategories where the multi-sigmoid models the development more accurately without overfitting.

access distribution service, created in 1991, of more than 2 million scholarly articles related to more than 170 technical domains. The metadata and scientific texts are openly available for download and comprise more than 3 TB of `.pdf` files [9]. The metadata contains information on each uploaded e-print, including authors, title, associated subcategory, and publication date. The `arXiv` computer-science category distinguishes between 40 different subcategories according to the ACM Computing Classification System [10]. We analyze the same `arXiv` subcategories that Percia et al. examined. Table 1 gives an overview of the selected subcategories. The primary metric we use in this work to quantify technological development is the number of e-print uploads to `arXiv` as a proxy for the development of the technologies associated with the subcategory in which the e-print was uploaded in. We calculate this metric with a monthly frequency to balance the velocity and the amount of data.

Sutton and Gong [11] looked at the computer-science subcategories on `arXiv` and found that the total fraction of e-prints published in top computer science conferences that are also available on `arXiv` rose from under 1% in 2007 to 23% in 2017. This increased attention on `arXiv` might skew the growth of some subcategories. To remove this bias of increased attention on `arXiv`, we normalize the monthly number of e-print uploads in a computer-science subcategory by the monthly total number of uploads on `arXiv`.

We hypothesize that the selected subcategories' growth might better be modeled by a multi-sigmoid model as described by Meyer [7]. Therefore, as our contribution, we set out to validate the following hypothesis:

H1: *The technological development of subcategories on* `arXiv` *follows a multi-sigmoid growth pattern.*

To validate the presented hypothesis, we fit a single-sigmoid and a multi-sigmoid to each of the 20 time series describing the number of uploaded e-prints in a computer-science subcategory. To calculate the fit, we use a non-linear optimization procedure and compare fit metrics to determine the quality of the fit. The solver we use requires lower and upper bounds for each fitted parameter (the single-sigmoid has four parameters and thus requires eight optimizer hyperparameters). Using a grid search, we find the hyperparameters that yield the best χ_v.

Out of the 20 subcategories analyzed, we determine that the development of nine subcategories is modeled more accurately by a multi-sigmoid, cf. Table 1. On top of that, we can replicate a single-sigmoid with a multi-sigmoid by setting the additional parameters to zero. This makes the multi-sigmoid the more general model, and thus, the more accurate one to describe technological development. However, the optimization algorithm tends to overfit the multi-sigmoid, so visual verification of the resulting fit is needed.

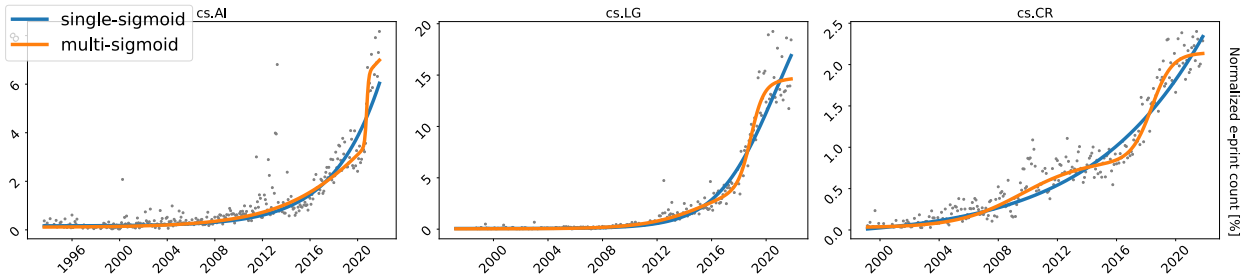

Figure 1: **Fits for the subcategories** `cs.AI`**,** `cs.LG`**, and** `cs.CR`. We can see that `cs.AI` and `cs.LG` that not yet reach their inflection point, and thus can be expected to grow in the future. The second growth spurt visible in `cs.CR` might be caused by the growth in AI and LG.

Furthermore, we find important insights for decision-makers: we discover a significant second growth phase for the subcategories "Information Retrieval", "Cryptography (CR)", and "Systems and Controls". The subcategories "Artificial Intelligence (AI)" and "Machine Learning (LG)" have not yet reached their inflection points and are thus expected to keep growing in the near future. We attribute the second growth spurt in the three previously mentioned subcategories to the recent solid growth in AI and LG. This means that especially the fields AI, LG, and CR have not yet reached their carrying capacity in research, i.e., we can expect further progress in those fields with significant impacts on industry.

References

(1) R. N. Stavins, A. B. Jaffe and R. G. Newell, *SSRN Electronic Journal*, 2002, DOI: 10.2139/ssrn.311023.

(2) S. Kingsland, *The Quarterly Review of Biology*, 1982, **57**, Publisher: University of Chicago Press, 29–52.

(3) D. Percia David, W. Lacube, S. Gillard, A. Mermoud, L. Maréchal, M. Tsesmelis and T. Maillart, *Security Dynamics in Computer Science Technologies*, Rochester, NY, 2022.

(4) J. Fisher and R. Pry, *Technological Forecasting and Social Change*, 1971, **3**, 75–88.

(5) A. Lotfi, A. Lotfi and W. E. Halal, *Technology Analysis & Strategic Management*, 2014, **26**, 943–957.

(6) D. Kucharavy and R. De Guio, *Procedia Engineering*, 2011, **9**, 402–416.

(7) P. S. Meyer and J. H. Ausubel, *Technological Forecasting and Social Change*, 1999, **61**, 209–214.

(8) P. Meyer, *Technological Forecasting and Social Change*, 1994, **47**, 89–102.

(9) arXiv, *arXiv Dataset*, 2020, https://www.kaggle.com/datasets/Cornell-University/arxiv (visited on 06/10/2022).

(10) ACM, *ACM Classification Codes*, 2020, https://cran.r-project.org/web/classifications/ACM.html (visited on 06/10/2022).

(11) C. Sutton and L. Gong, *Popularity of arXiv.org within Computer Science*, Number: arXiv:1710.05225, 2017.