Knowledge does not explode but increases linearly over time

Xinbing Wang (xwang8@sjtu.edu.cn)
Shanghai Jiao Tong University https://orcid.org/0000-0002-0357-8356

Huquan Kang
Shanghai Jiao Tong University

Luoyi Fu
Shanghai Jiao Tong University

Jiaxin Ding
Shanghai Jiao Tong University

Shiyu Liang
Shanghai Jiao Tong University

Jianghao Wang
Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences
https://orcid.org/0000-0001-5333-3827

Lei Zhou
Shanghai Jiao Tong University https://orcid.org/0000-0002-0433-3991

Chenghu Zhou
Institute of Geographic Sciences and Natural Resources Research https://orcid.org/0000-0003-3331-2302

Social Sciences - Article

Keywords:

Posted Date: July 7th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3102533/v1

Abstract

ChatGPT and GPT-4 have raised debates regarding the progress of knowledge in large language models [1-3]. The notion of "knowledge explosion" has been controversial in various variations since the 19th century [4-8]. Despite numerous indications to the contrary [9-11], conclusive evidence on knowledge growth is lacking [12]. Here, we evaluated knowledge as a collective thinking structure within citation networks by analyzing large-scale datasets containing 213 million publications (1800–2020) and 7.6 million patents (1976–2020). We found that knowledge did not explode but grew linearly over time in naturally formed citation networks that expanded exponentially. Our theoretical analysis established that the knowledge never exceeds the size of the network, revealing the limitation of knowledge development. Moreover, our results showed that the knowledge expansion rate shifted at certain inflection points, implying quantitative-driven qualitative changes. Leaps near inflection points may instigate a "knowledge explosion" delusion, allowing us to reconcile the spreading of the misconception. Inflection points in knowledge growth exhibited similar characteristics to the emergent ability of artificial intelligence [13], furnishing fresh insights into the singularities and emergence in complex systems. Overall, our findings reveal a slow pace of knowledge compared to data, reacquainting us with the progress of knowledge over time.

Main

Recently, ChatGPT, particularly GPT-4, has demonstrated remarkable capabilities [1], sparking a revival of explosive growth. The widely held belief that knowledge is growing exponentially worldwide raises this resonance, with rapid technological advancement and the proliferation of online information supporting this belief [14–18]. The concept of "knowledge explosion" was first introduced into academic literature as early as 1961, when the term was credited as having been coined [14]. Since then, the "knowledge explosion" has frequently surfaced in academic literature, enduring to the present day and spanning diverse fields such as education [15], politics [16], and scientific research [17,18]. Recent studies, however, indicate that the view of exponential knowledge growth may be overly simplistic [7–11]. For instance, the expansion of Wikipedia, which encompasses vast knowledge, has decelerated [10]. As early as 1962, Nobel laureate Fritz Albert Lippmann challenged this view and pointed out that despite claims of exponential knowledge growth, human evolution as a species does not change significantly [4]. Subsequently, numerous studies indicated that the rate of critical scientific discoveries declines [6,9] and that extensive literature may impede the creation of new knowledge [11]. All these studies indicate the same conclusion: knowledge may not be exploding as rapidly as previously believed.

Although the exact origin of the "knowledge explosion" is hard to trace from the literature, the concept of "information explosion" emerged at almost the same time [19], possibly leading to confusion between knowledge and information. "Information explosion" usually refers to the rapid increase in published information or data due to technological advances, such as television and computers. Born in the same era, "information explosion" and "knowledge explosion" describe similar and vague meanings, such that

later studies have used the two terms indiscriminately [20,21]. Studies have shown that the so-called "knowledge explosion" is just an "information explosion" [5]. Similarly, some studies have pointed to limitations in the accessibility and quality of the information available online, suggesting that the proliferation of information does not necessarily equate to increased knowledge [6,22,23]. Another related concept of "intelligence explosion," also known as the singularity, was first proposed in the 1950s [24] and gained popularity in the 1960s [25,26]. This idea suggests that technological advances will lead to artificial intelligence surpassing human intelligence, resulting in a singularity where the future becomes unpredictable. Though not entirely identical to "knowledge explosion," the close connection between knowledge and technology allows for mutual inspiration in exploring both. Recent research suggests that the exponential runaway technology trend favored by proponents of the singularity hypothesis has not been observed so far [27,28], which means that the "knowledge explosion" may not have occurred yet.

Is "knowledge explosion" occurring? This question has been raised, but no answer has been given due to the difficulty of measuring knowledge [12]. Some studies indicate that the network structure may potentially balance information explosion and knowledge mining [29]. Developing such an idea, the recently proposed knowledge quantification index (KQI) [30] provides a metric for knowledge by conceptualizing knowledge as a network structure, which embodies the certainty of information the structure brings. The intuition of KQI is that a well-structured network filled with knowledge should vastly reduce uncertainty about the unknown. Regarding each publication as an indivisible knowledge point and collective thinking structure constructed from citations as a knowledge network, KQI quantifies the uncertainty reduced by the knowledge network compared to isolated knowledge points. In this manner, KQI allows us to explore whether knowledge is exploding and how the knowledge grows.

Here, we addressed these confusions on knowledge growth by analyzing 213 million publications (1800–2020) in Microsoft Academic Graph (MAG) and 7.6 million patents (1976–2020) in the United States Patent and Trademark Office's (USPTO) Patents View database (see *Methods*). The MAG data include a field classification of the publications into 19 major disciplines and 292 secondary subjects. We constructed yearly citation network snapshots using MAG data, Patents View data, and subject data obtained by partitioning MAG. Subsequently, we joined the knowledge measure KQI [30] with analyses of each snapshot to observe changes in knowledge over time (see *Methods*). To comprehensively understand the findings, we also observed the changes in the duration of mathematical conjecture proofs over the past six decades and the changes in knowledge on random graph models. Using these data, we disclosed laws of knowledge growth and potential sources of the delusion of "knowledge explosion."

Results

Linear growth of knowledge

Although the concept of "knowledge explosion" has been popular for decades [14,15,17], we found that the later terminology referred more to "information explosion" than to "knowledge explosion," which may be a

misunderstanding. The Google Scholar search results for "knowledge explosion" and "information explosion" (see *Methods*) showed that, since 1977, the results of "information explosion" have outpaced those of "knowledge explosion," and this gap has more than doubled since 1987 and more than tripled since 2008 (Fig. 1a,d). In recent years, contrasted with the soaring results related to "information explosion," the results related to "knowledge explosion" seemed to have stagnated. These statistics suggested that academic communities were increasingly inclined to acknowledge "information explosion" rather than "knowledge explosion."

We found similar phenomena in the growth of KQI in publication and patent data. The number of publications has been increasing exponentially (Fig. 1e), in line with the idea of "information explosion." When analyzing the amount of knowledge, however, we found that it was not increasing at the same rate. Instead, the growth of KQI was almost linear (Fig. 1b), indicating that although there was a lot more information, people were not necessarily acquiring more knowledge [6,22,23]. Unlike the explosive growth in the number of publications, the number of patents was relatively slower (Fig. 1e). Remarkably, despite the significant difference between the growth rate of the number of patents and publications, the KQI of patents showed the same linear growth pattern (Fig. 1b). To verify this finding in more realistic scenarios, we split the academic publications by disciplines and subjects. Although the rate of knowledge growth varied across different fields, they all increased linearly starting from a certain year (Fig. 1c,f). The increase in knowledge was wildly out of step with scientific productivity and was not entirely parallel to the growth pattern of scientific productivity.

The linear growth pattern in knowledge is not limited to changes in KQI. As the foundation of natural sciences, mathematics represents the cutting edge of human knowledge at the time [31,32]. The formulation and proof of mathematical conjectures reflect the level of understanding and mastery of knowledge. The duration from formulation to proof corresponds to the time for knowledge to move from understanding to mastery. Over the past sixty years, no significant change occurred, indicating that human knowledge has constantly increased (Fig. 2, see *Methods*).

Numerous random graph models have been designed to capture the essential characteristics of real-world networks. Three commonly used models are the Erdős–Rényi (ER) model [33], the Barabási–Albert (BA) model [34], and the Watts–Strogatz (WS) model [35]. We simulated the growth patterns of knowledge in each of these models. The ER model is a simple random graph model where each node pair connects with a probability p. The BA model is a preferential attachment model where new nodes are added to the network one at a time and connected to existing nodes with a probability proportional to their degree. The results showed that the knowledge in both the BA and ER models increased linearly with the exponential growth of the number of nodes in the graph (Fig. 3a,b, see *Methods*). The WS model aimed to create complex networks between regular and random graphs by adjusting the rewiring probability. The results showed that knowledge growth in graphs that tend towards regularity was much faster than linear (Fig. 3c, see *Methods*). This result indicates that the linear growth of KQI is not confined to the design of KQI but rather a universal pattern embedded in real networks. KQI can grow at a striking speed in a highly regular simulated graph, but a limitation exists. This limitation states an inviolable law that knowledge

never grows beyond the size of the graph (see *Methods*). In other words, to keep the linear growth of knowledge, the size of the knowledge network must keep increasing at least linearly. The pattern of patent development closes to this boundary of simultaneous linear growth in both knowledge and graph size (Fig. 1b,e).

Inflection points in knowledge growth

Our examination of 19 major disciplines and 292 secondary subjects showed that knowledge tended to follow a linear growth pattern in most cases. However, there were significant inflection points between different stages of linear growth (Fig. 4b-e, Supplementary Fig. 1). Before the inflection point, the number of publications increased linearly at a specific rate and changed to another rate after the inflection point. Interestingly, there was nothing unusual about the number of publications around the inflection point. This phenomenon is similar to phase transition in complex networks. Understanding the regularity of these phase transitions is helpful for policymakers and researchers. Coreness is a measure that identifies tightly interlinked groups within a network [36], and we used mean coreness to characterize the overall connectivity of the network by averaging the corenesses of all the nodes. We found that when the network had a low mean coreness, meaning the academic network comprised numerous individual viewpoints, the probability of inflection points occurring was high (Fig. 4a). Conversely, when the mean coreness was high, meaning the academic network comprised tightly formed groups, the probability of inflections points occurring decreased. This finding suggests that academic diversity lays the foundation for knowledge innovation, necessitating tolerating heterodox viewpoints.

Diminishing returns of knowledge

The Pareto principle, also known as the 80/20 rule, describes the general imbalance of wealth distribution among individuals. We found a similar but even more imbalanced knowledge distribution in academia, where 9.27% of scientific productivity occupies 90.73% of knowledge; in comparison, 90.73% of scientific productivity occupies 9.27% of knowledge (Fig. 5a). In other words, a minimal increase in marginal returns of knowledge stems from an abundance of trivial literature as a backdrop, identifying with the stagnation coming from ossification of canon caused by massive literature [11]. This imbalance also implies that an increase in marginal returns of knowledge begins to diminish after establishing a certain level of background knowledge (Fig. 5b-e), consistent with the disruptiveness decline [9]. Similarly, the concept of "complexity brake" [8] manifests when the build-up of literature translates into a plateauing of knowledge gains. With each new area of knowledge augmentation becoming more complicated than the last, progress reaches a level wherein every new addition becomes almost trivial compared to what was there before. At this point, only nominal returns yield without a significant investment of time and resources. Despite the disproportion between investments and outcomes, it is necessary to expand scientific productivity for the little progress of knowledge if nothing can change the paradigm.

Discussion

In this study, we present findings that clarify the patterns of knowledge growth and explain the origins of "knowledge explosion" for the first time. Our results suggest that a linear model better describes knowledge growth, which is replicable using random graph models. This linear trend in knowledge growth has implications for predicting the emergence of new scientific breakthroughs, which may occur at more predictable intervals than previously assumed. The faster growth rate in the number of publications than knowledge growth implies that access to scientific research may become more widely available. However, this also raises concerns about academic inflation and the quality of research. These findings contribute to a deeper understanding of the dynamics of knowledge growth and its potential impact on the scientific community.

Whether knowledge is experiencing an explosion has been questioned in previous studies [4–6,12]. Quantifying knowledge, however, was intricate before introducing KQI as a metric [12]. Our results immaculately validate the previous skepticism that knowledge did not grow exponentially. Considering that the concepts of "knowledge explosion" and "information explosion" emerged almost simultaneously, it is likely that the former is a misunderstanding of the latter. Recent studies that suggest the decreasing disruptiveness [9] and the slowed canonical progress [11] in scientific fields imply a slowdown in knowledge growth, yet leave unanswered how much this slowdown occurs. Our findings, on the one hand, discard earlier optimistic claims of "knowledge explosion" and, on the other hand, clarify recent pessimistic claims of a scientific slowdown. Despite the slowdown of knowledge growth relative to the proliferation of publications, overall knowledge remains steadily increasing. Our concern in slowing scientific activities should be more on whether the expansion of science is sustainable and how to improve scientific efficiency.

The concept of singularity comes from the idea that technological advancements will lead to an exponential increase in knowledge, resulting in artificial intelligence surpassing human intelligence. However, if knowledge is not growing exponentially, the pursuit of intelligent singularity must be reevaluated because it may not be achievable soon. Inflection points in knowledge explain the singularity theory and "knowledge explosion" phenomenon, i.e., the sudden disparity between pre- and post-transition knowledge growth rates creates an illusion of exponential growth. These transition points suggest that fundamental changes occur, like knowledge acquisition, as scientific research reaches a certain level of expertise. Understanding this shift could provide valuable insights into the process of scientific discovery. Based on our preliminary observation of the distribution pattern of inflection points, it is easier to observe an inflection when the mean coreness is relatively small when many small research groups scattered throughout the network exist. This finding explains the previous discoveries that small teams promote knowledge innovation [37–39].

Some studies have shown inequality in academia [40–43]. Our reported inequality in knowledge contributions further reveals that the continuously expanding citation network causes such inequality. The accompanying diminishing returns of knowledge have similarities to the concept of the complexity brake [8], which postulates that the more we acquire knowledge, the more cognizant we become of the

vastness of the unknown and the more necessary to reassess our earlier perceptions. This connection implies that the growth of knowledge is subject to similar constraints as other complex systems. For instance, the disparity between the increased number of parameters and the improvement in performance while pursuing large-scale artificial intelligence models is consistent with the disparity between network scale and knowledge growth [13]. This relation inspires us to use knowledge to measure the capacity of neural networks and apply the law of knowledge growth to improve model performance. Unlike natural networks, neural networks are readily adjustable, providing an opportunity to approach the limit of knowledge growth speed, as seen in the WS model (Fig. 3c). The recent small yet high-performing models, such as LLaMA [44], corroborate this possibility. Notably, the inflection point in knowledge growth reflects the emergence of the capabilities in large models such as ChatGPT and GPT-4 [1,13]. These connections offer valuable insights for the further advancement of artificial intelligence.

The present study is not without limitations. The KQI is a novel metric for quantifying knowledge and has yet to gain widespread acceptance within the scientific community. Further investigation into the characteristics and value of KQI will contribute to its gradual recognition. Additionally, as books remain the primary means of disseminating knowledge, our research focuses primarily on academic publications and patents, limiting our analysis's scope. Future work involving various media types may deepen our understanding of knowledge evolution. As knowledge network construction is complicated, particularly in defining concepts and relationships, the present study utilized a citation network constructed using publications or patents as nodes without consideration for semantics. In the future, incorporating more diverse knowledge network types, such as knowledge graphs, may further enrich the results of analyzing knowledge evolution patterns.

In conclusion, the results of this study have significant consequences for comprehending the growth of knowledge. Using KQI as a metric for measuring knowledge growth provides a valuable tool. Moreover, our findings support emergence theories arising from inflection points in knowledge growth and the complexity brake. Further investigation is needed into the causes of the inflection points and the constraints on knowledge growth to gain a deeper understanding of this complex phenomenon. It also indicates that we should not be intimidated by the explosion of knowledge but rather cope with the exhaustion of the explosion of information. Despite the explosion in scientific productivity, calm down; we are only walking on the trail of exploring knowledge.

Methods

Publications data. Our data is derived from the MAG data, which archives publications from 1800 to 2021. The publications cover 292 secondary subjects in 19 major disciplines, including but not limited to Economics, Biology, Computer science, and Physics. We excluded patents, datasets, and repositories, utilizing the doctype field in the MAG data. We limited our focus to publications up to 2020 because recent literature was probably not sufficiently collected. Although we used literature from as early as 1800, the KQI was only calculated from 1920 because the citations were too sparse to be interconnected in the early years. We removed possible errors in the data, including self-citations, duplicate citations, and

citations violating time order. After eliminating potentially incorrect publications and closing the data up to 2020, the analytical sample consisted of 213,715,816 publications and 1,762,008,545 citations. The subject data is split from the MAG data. While processing data from a particular subject, we only preserved citation relationships that both article and reference are on the same subject, thus guaranteeing that all nodes within the network are from the same subject.

Patents data. The Patents View data collect 8.1 million patents granted between 1976 and 2022 and their corresponding 126 million citations. We limited our focus to citations made to U.S. granted patents by U.S. patents up to 2020 because recent patents were probably not yet sufficiently collected. Although we used patents from 1976, the KQI was only calculated from 2000 because the citations were too sparse to be interconnected in the early years. We removed possible errors in the data, including self-citations, duplicate citations, and citations violating time order. After eliminating potentially incorrect patents and closing the data up to 2020, the analytical sample consisted of 7,627,229 patents and 101,148,606 citations.

KQI calculation. KQI [30] is a metric that quantifies knowledge from the perspective of information structurization. As described by the proposers of KQI, we constructed year-by-year publication citation graphs or patent citation graphs, which are directed acyclic graphs. We calculated the KQI of each node in a citation graph and added them up to obtain the KQI of the citation graph for each year, exploiting the additivity of the KQI as pointed out by the proposers.

Search results in Google Scholar. The two terms, "knowledge explosion" and "information explosion," are searched as phrases enclosed with quotation marks in Google Scholar (https://scholar.google.com) and filtered by year ranges. We recorded the number of results returned manually.

Analysis of mathematical conjecture. We collected 61 mathematical conjectures proven to be correct since 1960 (Supplementary Table 1). Conjectures proved wrong and those not yet proven are excluded. We chose a possible intermediate year for conjectures without a specific formulation year or proof year. Due to certain mathematical conjectures proved in the same year but with different durations and difficulty in determining their temporal order, we took the average duration of proofs within the same year. We used such a time series when conducting correlation analyses and hypothesis testing. Spearman and Kendall [45] rank correlation coefficients are non-parametric measures of the strength of monotonic association between two variables and are calculated by measuring the rank correlation between two variables. The range of these two coefficients is from − 1 to 1, with values closer to 0 indicating a weaker relationship between the two variables. Cox-Stuart [46] and Mann-Kendall [47] hypothesis tests assess whether there is a monotonic increasing or decreasing trend over time in a time series data. The null hypothesis for both hypothesis tests is the absence of a monotonic trend, so a p-value greater than 0.05 indicates the lack of a significant trend.

Random graph generated by Barabási-Albert model. The BA model [34] uses a preferential attachment process to generate random graphs. We used the BA model under undirected graphs. After generating the

graph, we oriented each edge chronologically, from the node joined earlier to the node joined later. This method naturally resulted in directed acyclic graphs available for KQI calculation directly.

Random graph generated by Erdős-Rényi model. The ER model [33] generates a random graph with the same probability of existence for each edge between two arbitrary nodes. We used the ER model under undirected graphs and uniquely numbered each node. After generating the graph, we oriented each edge in numbered order, from the smaller numbered node to the larger numbered node. This method naturally resulted in directed acyclic graphs available for KQI calculation directly.

Random graph generated by Watts-Strogatz model. The WS model [35] generates graphs with small-world properties and is adjustable between regular and random graphs. Following the WS model, we constructed a regular ring lattice, rewired edges, and numbered each node incrementally and uniquely along the ring. After generating the graph, we oriented each edge in numbered order, from the smaller numbered node to the larger numbered node. This method naturally involved interpolating between a regular ring lattice and a random graph.

Limitation of knowledge growth. We prove that the growth rate of KQI has an upper bound regarding the graph size through a theoretical derivation from the KQI formula. We rewrite the formula of KQI: $K_\alpha = \sum_{\beta\to\alpha} \frac{d_\alpha^i V_\beta - V_\alpha}{d_\alpha^i W} \log\left(1 + \frac{1}{\frac{V_\alpha}{d_\alpha^i V_\beta - V_\alpha}}\right)^{\frac{V_\alpha}{d_\alpha^i V_\beta - V_\alpha}}$.Applying the Euler limit formula, it is simplified as follows:$K_\alpha = \frac{a}{W}\left(\sum_{\beta\to\alpha} V_\beta - V_\alpha\right), (0 < a < \log e)$.We sum KQI over all nodes and note that W is the sum of the out-degrees of all nodes. The relation between K and W is thus derived:
$K \triangleq \sum_\alpha K_\alpha = \frac{a}{W}\sum_\alpha\left(\sum_{\beta\to\alpha} V_\beta - V_\alpha\right) = a\sum_\alpha \frac{(d_\alpha^o - 1)}{W} V_\alpha < a\mathbb{E}\left(V_\alpha\right) < W\log e.$

Discovery of inflection points. We discover the inflection points using segmented regression models developed by Vito M. R. Muggeo [48]. The segmented regression was performed on the curve of KQI over time, with the regression line breakpoint considered the inflection point. We started with the null hypothesis of no breakpoint and performed a score test to determine if there was an additional breakpoint [49]. This process repeated until no additional breakpoint. The significance level was 0.01. To counteract the multiple comparisons problem, we employed Bonferroni correction, requiring that the p-values for each of the first k tests be smaller than 0.01/k. Once the number of breakpoints was determined, we used the segmented method (33) to estimate their positions.

Estimation of inflection density. The inflection density is a quantity that characterizes the distribution of the network state at its transition between two different regimes. The area under an inflection density curve represents the average times finding the system in the inflection state. The main text investigates the inflection density per unit mean coreness. Due to the estimation error of inflection points, we map the probability densities of normal distributions centered around the inflection points to the mean coreness using linear interpolation and summing in cases with multiple mapping values. The estimated standard

deviation of the inflection point determines the standard deviation of the normal distribution. We estimated the inflection density of a discipline by summing the probability densities with respect to mean coreness during network evolution. The total inflection density is the mean of densities for all disciplines. To estimate confidence intervals, 1000 bootstrap resamplings are employed.

Calculation of marginal KQI. The marginal KQI is calculated by subtracting the KQI for a given graph of the previous year from the current year and dividing it by the number of nodes added in the current year. We applied the LOWESS (locally weighted scatterplot smoothing) nonparametric regression method to perform local regression of marginal KQI. To estimate the 95% confidence interval of the LOWESS fit, we performed 1000 bootstrap resamplings. The fraction of data used when estimating was 2/3.

Declarations

Data availability

Data from MAG and Patents View are publicly available. Data from MAG are requested from Acemap (https://www.acemap.info/) at Shanghai Jiao Tong University and are available at https://zenodo.org/record/7878551. Data from Patents View are available at https://patentsview.org/. Other source data are provided with this paper.

Code availability

Open-source code for calculating KQI is available at https://github.com/Girafboy/KQI.

Acknowledgments

This work was supported by NSF China (No. 42050105, 62020106005, 62061146002, 61960206002), Shanghai Pilot Program for Basic Research - Shanghai Jiao Tong University. Thanks to Yuang Ding for collecting the mathematical conjecture data.

Author Contributions

H. K., X. W., and L. F. collaboratively contributed to the conception and design of the study. X. W. contributed to the acquisition of the data. H. K. created software used in the study and drafted the manuscript. X. W., L. F., J. D., S. L., J. W., L. Z., and C. Z. collaboratively revised the manuscript.

Competing interests.

The authors declare no competing interests.

Supplementary information

Supplementary Information. Supplementary Tables 1–2 and Figure 1.

References

1. Bubeck, S. *et al.* Sparks of Artificial General Intelligence: Early experiments with GPT-4. (2023).
2. Ienca, M. Don't pause giant AI for the wrong reasons. Nature Machine Intelligence 5, 470–471, doi:10.1038/s42256-023-00649-x (2023).
3. Clarke, L. Call for AI pause highlights potential dangers. Science (New York, NY) 380, 120–121 (2023).
4. Lipmann, F. Disproportions created by the exponential growth of knowledge. Perspectives in biology and medicine 5, 324–326 (1962).
5. Eleuterio, H. Macromolecular scientists: From pot-boilers to programmers. Journal of Chemical Education 53, 352 (1976).
6. Horgan, J. The End of Science: Facing the Limits of Knowledge in the Twilight of the Scientific Age. (1996).
7. Walsh, T. The singularity may never be near. ai Magazine 38, 58–62 (2017).
8. Allen, P. & Greaves, M. The singularity isn't near. Technology review 12, 7–8 (2011).
9. Park, M., Leahey, E. & Funk, R. J. Papers and patents are becoming less disruptive over time. Nature 613, 138–144, doi:10.1038/s41586-022-05543-x (2023).
10. Suh, B., Convertino, G., Chi, E. H. & Pirolli, P. The singularity is not near: slowing growth of Wikipedia. *Proceedings of the 5th international symposium on wikis and open collaboration*, 1–10 (2009).
11. Chu, J. S. & Evans, J. A. Slowed canonical progress in large fields of science. *Proceedings of the National Academy of Sciences* 118 (2021).
12. Ehrlich, P. R. *et al.* Knowledge and the environment. Ecological economics 30, 267–284 (1999).
13. Wei, J. *et al.* Emergent Abilities of Large Language Models. ArXiv abs/2206.07682 (2022).
14. Swift, E. H. Keeping the Curriculum Up to Date. Engineering and Science 25, 11–15 (1961).
15. Tachibana, T. Closing the knowledge gap between scientist and nonscientist. Science 281, 778–779 (1998).
16. Galas, D. J. & Riggs, H. Global Science and U.S. Security. Science 300, 1847–1847, doi:doi:10.1126/science.300.5627.1847 (2003).
17. Adair, J. G. & Vohra, N. The explosion of knowledge, references, and citations: Psychology's unique response to a crisis. American Psychologist 58, 15 (2003).
18. Prasad, M., Freitas, H., Fraenzle, S., Wuenschmann, S. & Markert, B. Knowledge explosion in phytotechnologies for environmental solutions. Environmental Pollution 158, 18–23 (2010).
19. ZAND, D. E. The Information Explosion. *Academy of Management Proceedings* 1961, 44–46 (1961).
20. Mahajan, R., Gupta, P. & Singh, T. Massive open online courses: concept and implications. Indian pediatrics 56, 489–495 (2019).
21. Zhang, X. On the Application of Computer Network Information Technology in Educational Reform. (2020).

22. Chen, Y. & Xu, D. Computational analyses of high-throughput protein-protein interaction data. Current protein and peptide science 4, 159–180 (2003).
23. Weiss, P. A. Whither Life Science? What are the great unanswered questions that serve as beacons for biological research? American Scientist 58, 156–163 (1970).
24. Ulam, S. Tribute to John von Neumann. Bulletin of the American mathematical society 64, 1–49 (1958).
25. Good, I. J. *Speculations concerning the first ultraintelligent machine*. Vol. 6 (Elsevier, 1966).
26. Vinge, V. Technological singularity. *VISION-21 Symposium sponsored by NASA Lewis Research Center and the Ohio Aerospace Institute*, 30–31 (1993).
27. Eden, A., Moor, J., Søraker, J. & Steinhart, E. *Singularity Hypotheses: A Scientific and Philosophical Assessment*. (Springer, 2013).
28. Modis, T. Links between entropy, complexity, and the technological singularity. Technological Forecasting and Social Change 176, 121457 (2022).
29. Krishna, A. Web 2.0 and the ever elusive balance between information explosion and data mining. Nature Precedings, doi:10.1038/npre.2008.1959.1 (2008).
30. Wang, X. *et al.* Quantifying knowledge from the perspective of information structurization. Plos One 18, e0279314, doi:10.1371/journal.pone.0279314 (2023).
31. Wigner, E. P. The unreasonable effectiveness of mathematics in the natural sciences. Richard courant lecture in mathematical sciences delivered at New York University, May 11, 1959. *Communications on Pure and Applied Mathematics* 13, 1–14, doi:https://doi.org/10.1002/cpa.3160130102 (1960).
32. Wigner, E. P. in *Mathematics and science* 291–306 (World Scientific, 1990).
33. Erdős, P. & Rényi, A. On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci 5, 17–60 (1960).
34. Barabási, A.-L. & Albert, R. Emergence of Scaling in Random Networks. Science 286, 509–512, doi:doi:10.1126/science.286.5439.509 (1999).
35. Watts, D. J. & Strogatz, S. H. Collective dynamics of ‘small-world’ networks. Nature 393, 440–442, doi:10.1038/30918 (1998).
36. Seidman, S. B. Network structure and minimum degree. Social Networks 5, 269–287, doi:https://doi.org/10.1016/0378-8733(83)90028-X (1983).
37. Wu, L., Wang, D. & Evans, J. A. Large teams develop and small teams disrupt science and technology. Nature 566, 378–382 (2019).
38. Xu, F., Wu, L. & Evans, J. Flat teams drive scientific innovation. *Proceedings of the National Academy of Sciences* 119, e2200927119, doi:doi:10.1073/pnas.2200927119 (2022).
39. Zeng, A., Fan, Y., Di, Z., Wang, Y. & Havlin, S. Fresh teams are associated with original and multidisciplinary research. Nature Human Behaviour 5, 1314–1322, doi:10.1038/s41562-021-01084-x (2021).

40. Wapman, K. H., Zhang, S., Clauset, A. & Larremore, D. B. Quantifying hierarchy and dynamics in US faculty hiring and retention. Nature 610, 120–127, doi:10.1038/s41586-022-05222-x (2022).
41. Liu, F., Holme, P., Chiesa, M., AlShebli, B. & Rahwan, T. Gender inequality and self-publication are common among academic editors. Nature Human Behaviour 7, 353–364, doi:10.1038/s41562-022-01498-1 (2023).
42. Li, W., Zhang, S., Zheng, Z., Cranmer, S. J. & Clauset, A. Untangling the network effects of productivity and prominence among scientists. Nature Communications 13, 4907, doi:10.1038/s41467-022-32604-6 (2022).
43. Gomez, C. J., Herman, A. C. & Parigi, P. Leading countries in global science increasingly receive more citations than other countries doing similar research. Nature Human Behaviour 6, 919–929, doi:10.1038/s41562-022-01351-5 (2022).
44. Touvron, H. *et al.* Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).
45. Kendall, M. G. A new measure of rank correlation. Biometrika 30, 81–93 (1938).
46. Some quick sign tests for trend in location and dispersion. Biometrika 42, 80–95 (1955).
47. Mann, H. B. Nonparametric tests against trend. Econometrica: Journal of the econometric society, 245–259 (1945).
48. Muggeo, V. M. Segmented: an R package to fit regression models with broken-line relationships. R news 8, 20–25 (2008).
49. Muggeo, V. M. Testing with a nuisance parameter present only under the alternative: a score-based approach with application to segmented modelling. Journal of Statistical Computation and Simulation 86, 3059–3067 (2016).

Figures

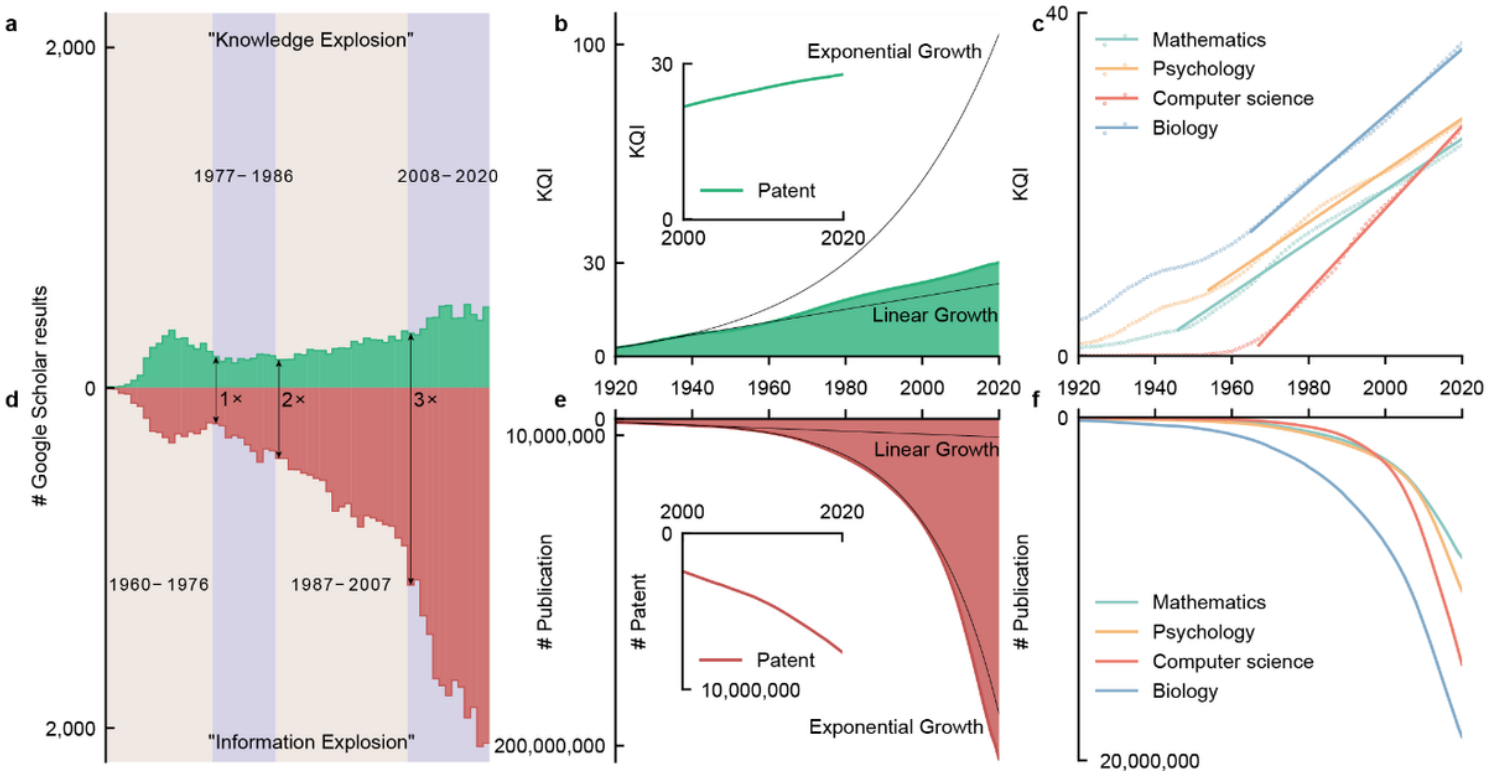

Figure 1

Linear growth of knowledge. a and d, The number of results obtained by searching for "knowledge explosion" (**a)**and "information explosion" (**d**) in Google Scholar year by year from 1960 to 2020. From 1960 to 1976, "information explosion" results are roughly the same as "knowledge explosion" results. From 1977 onwards, "information explosion" results are more than "knowledge explosion" results; this gap has been more than twice since 1987 and more than tripled since 2008. **b** and **e**, Growth of KQI (**b)**and the number of publications or patents (**e**) in MAG (main plot) and Patents View (inset plot). The black curves provide referents for linear and exponential growth trends. **c** and **f**, Growth of KQI (**c**) and number of publications (**f**) in the disciplines of mathematics (green), psychology (orange), computer science (red), and biology (blue). Straight lines exhibit trends approaching linearity starting from certain years.

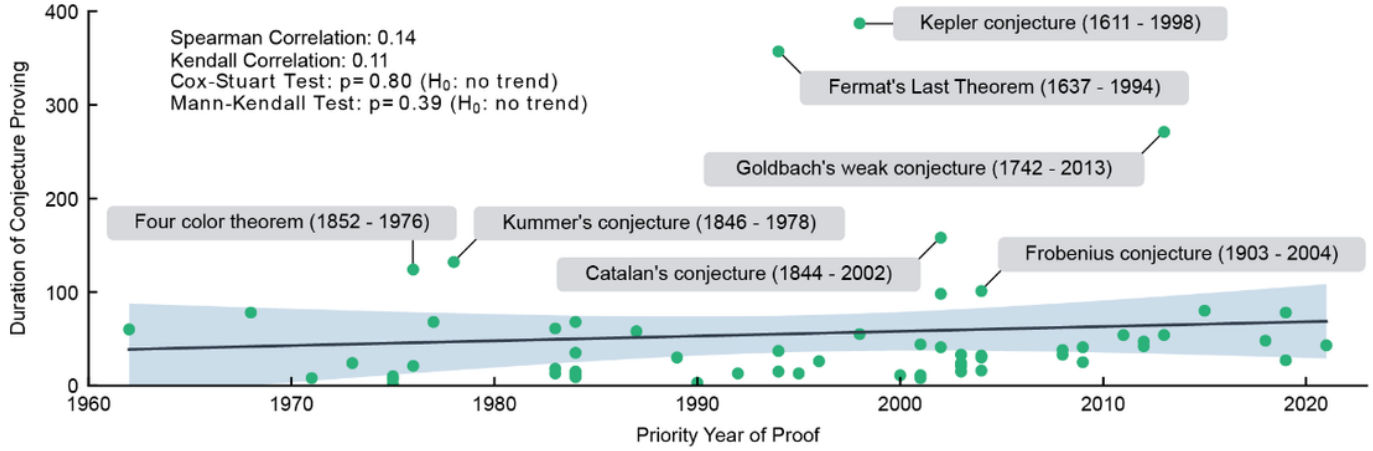

Figure 2

Duration of mathematical conjecture proving. The green scatter shows the duration (from the formulation to the proof completion, in years) of mathematical conjectures proved since 1960, with several notable examples highlighted. The solid black line is the least square linear regression, and the blue shaded band represents the 95% confidence interval. Spearman and Kendall rank correlation coefficients indicate a weak relationship between the duration of conjecture proving and the priority year of proof. The Cox-Stuart and Mann-Kendall hypothesis tests show that the duration of the mathematical conjectures' proofs has not changed significantly.

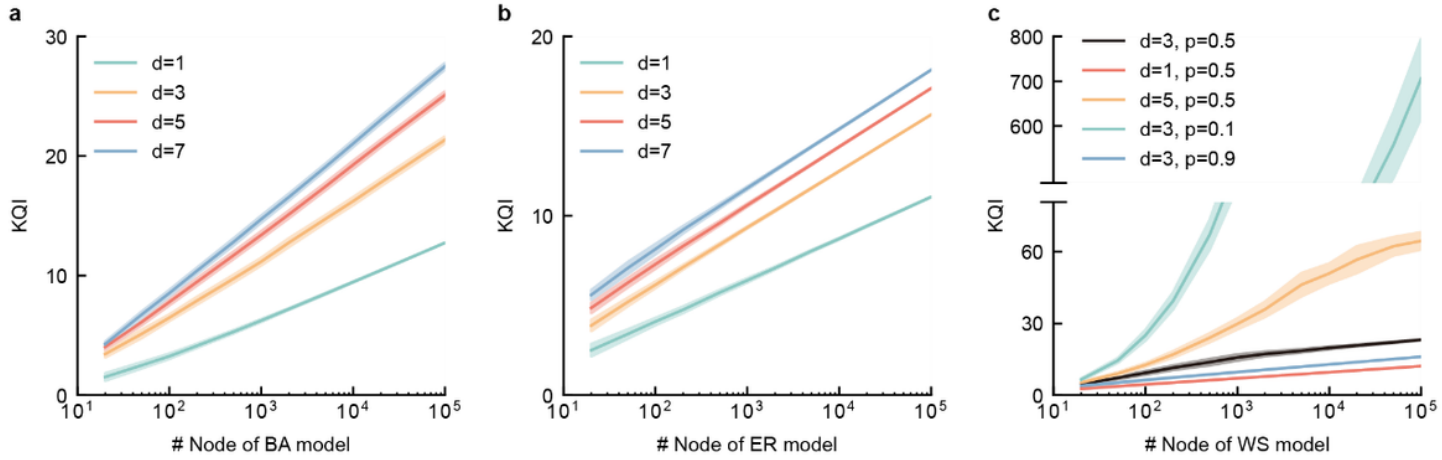

Figure 3

Knowledge growth on random graph models. a, Simulated KQI on graphs generated from the BA model. Each newly arrived node connects to *d* previous nodes in the BA model. Shaded bands mark the standard deviation. **b**, Simulated KQI on graphs generated from the ER model. Each node has, on average, *d* out-edges and *d* in-edges randomly connected to the other *n*-1 nodes. Shaded bands mark the standard deviation. **c**, Simulated KQI on graphs generated from the WS model. In a ring of *n* nodes connected to its 2*d* nearest neighbors, each edge is reconnected with probability *p* to another node uniformly at random. The vertical axis is truncated into two parts with different scales to show the details of the significantly different results. Shaded bands mark the standard deviation.

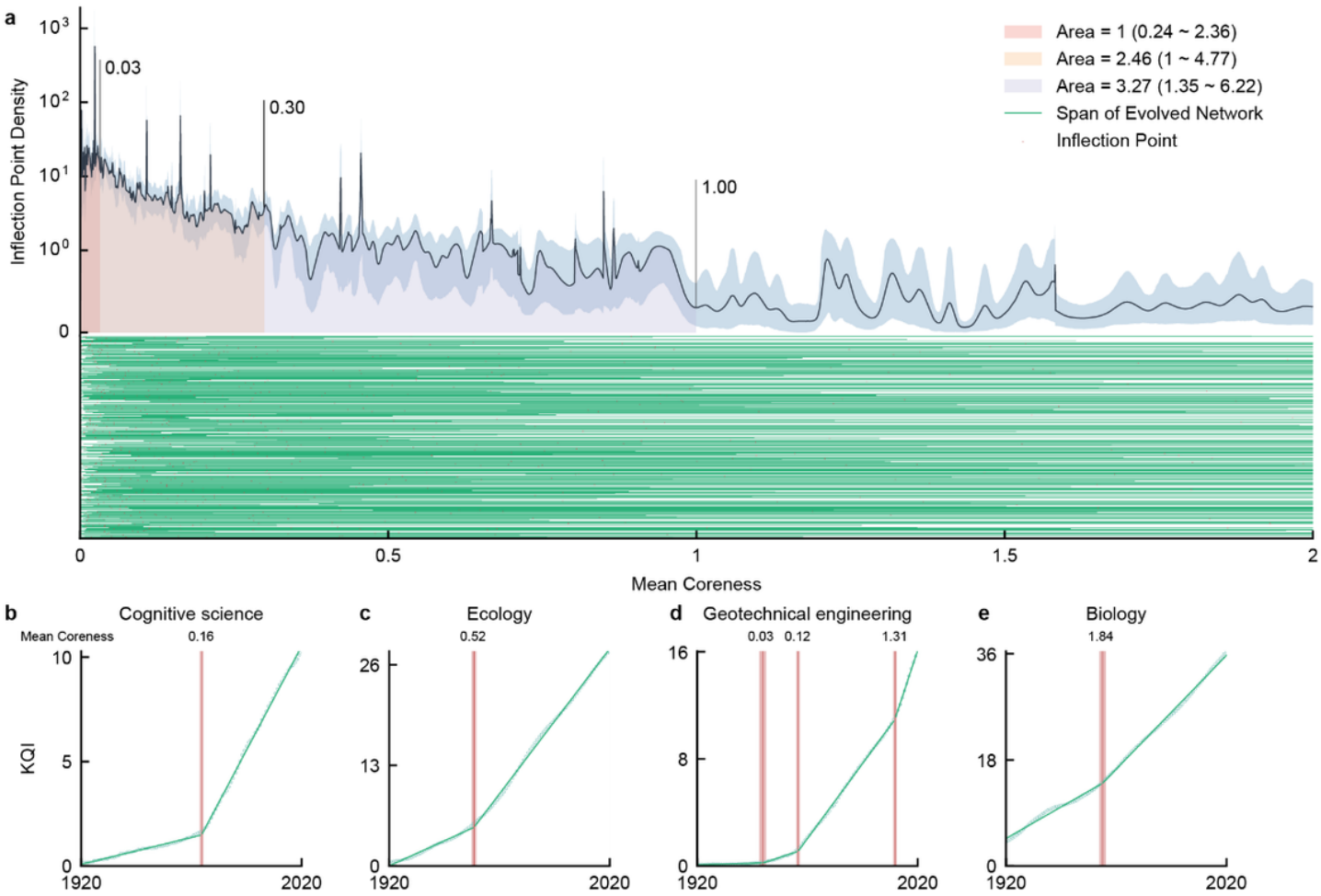

Figure 4

Inflection points in KQI evolution. a, Distribution of inflection points with mean coreness. Each green line represents evolving network of a discipline, and its extent on the x-axis corresponds to the range of mean coreness for the evolving network. The red point indicates the occurrence of an inflection point at this location of the evolving network. The inflection density (see *Methods*) is plotted as a black line, while a blue-shaded band indicates the 95% confidence interval. The y-axis is scaled using symlog, producing a linear plot within the specified range of values near zero (<1). Three regions of interest in the density curve are highlighted using different colored shaded regions. During the evolution of the network, on average, one inflection point occurs as the mean coreness increases from 0 to 0.03. There is a 95% probability of experiencing at least one inflection as the mean coreness increases from 0 to 0.30. During the increase in the mean coreness from 0 to 1, an average of 3.27 inflection points occur. **b-e**, Inflection points in distinct disciplines. The green circles represent the KQIs computed for the entire network at different years. The green lines are the segmented linear regression results. The red lines denote the estimated inflection points, and the red shaded bands represent the standard deviations. The mean corenesses of the network at inflection points are also marked.

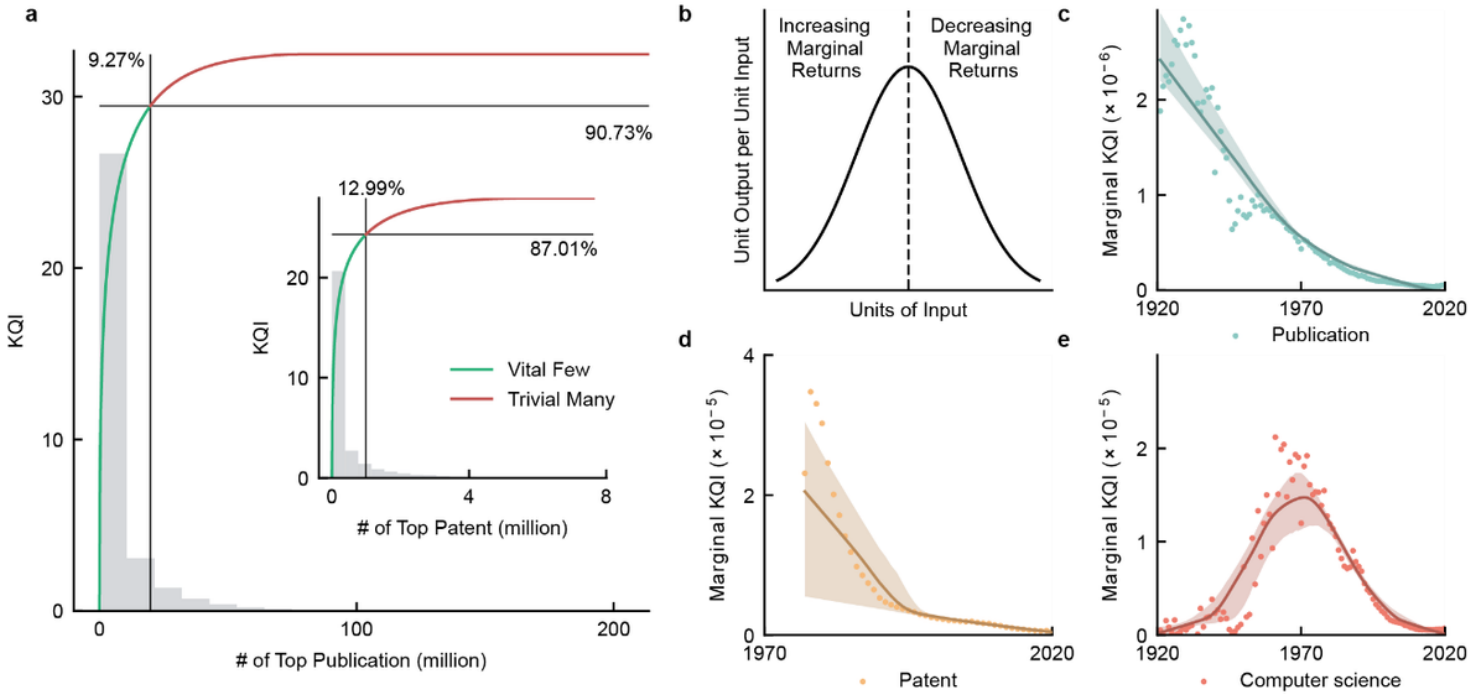

Figure 5

Pareto principle and diminishing returns in KQI. a, Cumulative KQI distribution after sorting publications by descending KQI order. The grey bar represents a histogram of the publications sorted in descending order according to KQI with KQI as its weight, expressing the contribution of the respective range of publications on KQI. The green and red lines are the cumulative curves of the KQI distribution; at their critical point, the ratio of publications in the vital few equals the ratio of trivial many to the total KQI contribution. The main plot displays statistics for publications in MAG, while the inset plot displays statistics for patents in Patents View. **b**, Demo illustrating the law of diminishing returns. **c-e**, Marginal KQI (see *Methods*) increment per publication (or patent) over time. The circle markers represent the average incremental KQI from each publication (or patent) in that year. The solid curves show the local regression, and the shaded bands indicate the 95% confidence interval.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- natureSIv7.2.pdf