Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-026-06490-y OPEN

Deepening and broadening knowledge after the PISA scientific event: bibliometric, semantic network, and expert analyses of scientization in education research

David P Baker[1], Abdul Basit Adeel[1✉], Juan José Moradel-Vásquez[2], Bea Treena Macasaet[2], Justin J. W. Powell[2], Julian Prieto[3], Jennifer Dusdal[2], Jun Pang[2] & Yuan Chih Fu[4]

The intensification of science as a social institution—known as scientization—is a hallmark of post-industrial society, characterized by deepening research within primary fields and expanding into new ones. While computational analyses of large bibliometric datasets reveal historical trends in the growth of science, their scale misses underlying dynamics. Conversely, qualitative accounts of specific discoveries provide context but fail to capture broader processes driving scientization. This paper introduces a middle-range approach to studying the expansion of science, focusing on scientific responses to significant scientific events ("S-events") within specific fields over fixed periods. By combining bibliometric data with natural language processing analyses of semantic networks and expert assessments, this approach sheds light on the development of epistemic communities and their role in deepening, broadening, and interacting to foster scientific progress. The Programme for International Student Assessment (PISA) event generated discovery by facilitating a fluid epistemic community of interconnected scientific papers, journals, and scientists. The accumulation of PISA papers appeared not only in directly relevant journals, but also in those at cognitive distance from educational learning science. A dynamic of new ideas broadened the main conceptual core network, while other ideas formed short-lived distal conceptual cores. This case study highlights the strengths and limitations of this middle-range approach for investigating scientization processes.

[1] Pennsylvania State University, State College, PA, USA. [2] University of Luxembourg, Esch-sur-Alzette, Luxembourg. [3] University of Toronto, Toronto, ON, Canada. [4] National Taipei University of Technology, Taipei, Taiwan. ✉email: axa6372@psu.edu

Introduction

Over the past century and a quarter, science, as a social institution, has developed an unprecedented capacity for discovery (Gauchat, 2023; Merton, 1996). Globally, research is conducted by scientists worldwide affiliated with ~41,000 universities of various types and an increasing number of non-academic organizations (Baker and Powell, 2024). The institution's discovery potential, reflected in the publishing of millions of scientific articles annually, has doubled about every 15 years from 1900 to 2015, and has nearly doubled again since then (Bornmann, Haunschild and Mutz, 2021). Growth across the "century of science" (Powell et al., 2017) was paralleled by an almost identical exponential increase in the number of research scientists up to a current estimated nine million worldwide (Dong et al., 2017; Zapp 2022). The extension of the scientific method across more topics and concepts is evident from the proliferation of journals and ongoing specialization. Reflecting the expansion of scientific topics, from 1950 to 1980, mainstream STEMM journals expanded by over 700% and by an additional 110% up to 2010 to about 46,000 contemporary journals (excluding predatory journals), including an estimated 9000 of the leading, indexed, and most cited in their fields (Web of Science Group, 2022).

The term scientization is increasingly used to denote both, the expansion of physical and human resources dedicated to research capacity, as well as the broader sociological process through which maturing social institutions become more internally differentiated and autonomous from other institutions (Marques et al., 2025). Empirically, scientization can be observed as the historical growth of the capacity to conduct research, and sociologically, as the concurrent consolidation of science into a fully developed social institution. Earlier studies have examined the influence of scientization on other institutional domains, such as the growing role of scientific knowledge in shaping policy and governance across diverse sectors (e.g., Drori et al., 2003; Schofer, 2003; Schofer and Hironaka, 2005). Yet, comparatively less attention has been paid to the internal development of scientization itself, including its evolving networks of researchers, modes of knowledge production, and conceptual frameworks. Although science is often assumed to be self-perpetuating—where past discoveries disrupt established paradigms and inspire new trajectories—internal mechanisms driving scientization remain underexplored (Baker et al., 2025).

Also motivating the present study is the criticism that amassing scientization may lead to an inflation of research output accompanied by fewer truly "disruptive" discoveries (e.g., Bloom et al., 2020; Landhuis, 2016; Hanson et al. 2024; Horgan, 2018; Park, Leahey and Funk, 2023; Milojević, 2015). Such critiques arise from interpretations of large bibliometric datasets that aggregate across all scientific domains and specialties, thereby likely obscuring the internal heterogeneity and processes of scientific development. In contrast, a growing counterargument suggests that rather than signaling inflation or decline, scientization may be transforming the tempo and dynamics of discovery itself (Baker et al., 2025; Holst et al., 2024). Demonstrating this possibility, however, requires a more fine-grained and integrative empirical approach.

Complementary methods, if integrated, promise to advance a fuller understanding of scientization by enabling the identification of patterns in the internal organization of science over time. Building on past research on fronts and specialties, we develop and illustrate here a novel middle-range approach consisting of three facets (e.g., Fujita et al., 2014; Fuchs 1992; Michaelson 1993; Morris and Van der Veer Martens, 2008). First, the combining of computational analyses of bibliometric metadata of frequencies of authors and research articles (hereafter, "papers") with natural language processing and semantic network analyses of concepts found in published papers. Second, the tracing of bibliometric and conceptual responses after a notable scientific event (hereafter, "S-event") within a specific field over a fixed period. And third, the enlisting of judgements of scientific field experts in identifying the S-event, setting a period length, and interpreting indicators of the scientific response.

Studies applying such methods across extensive datasets and long temporal spans have illuminated broad structural trends in scientific development at various levels (e.g., Wang and Barabási, 2021; Krauss, 2024). Yet the very scale that makes these approaches powerful also limits their ability to capture the mechanisms and institutional dynamics resulting from scientization. Conversely, qualitative historical accounts of major discoveries provide rich contextual detail but often fall short of revealing the institutionalized processes underlying scientization (e.g., Collins, 2018). Integrating these quantitative and qualitative perspectives may therefore yield a more comprehensive view of how scientization unfolds and transforms the production of knowledge.

Of course, scientific events are always implied in the study of research fronts, what is developed here is a more explicit approach that, if used more widely, adds the possibility of comparing responses to different types of events. Suitable S-events include: empirical rejection of a theory leading to new conceptualization; a disruptive empirical finding, either expected or unexpected; an external shock or other pressing natural and social crises (e.g., pandemics, climate change, population pressure); and, introduction of a transformative research instrument. The latter type, specifically the PISA dataset, serves as an illustrative case here.

Seeking a broader empirical lens than the conventional focus on co-authorship networks as indicators of research fronts (Glänzel and Schubert, 2004), bibliometric indicators are developed about the frequency of researchers and papers, plus semantic networks of concepts and their papers engaging with networks of scientific concepts. This enhanced approach draws on the notion of a scientific epistemic community, encompassing researchers, papers, and concepts; a summative term now widely employed in the study of knowledge generation (e.g., Adler and Haas, 1992; Haas, 2015; Miller and Fox, 2001).

To demonstrate the advantages of this approach, we examine the degree to which three specific scientization processes occurred following a specific S-event over a specified time. The first process is deepening, reflected in the intensified growth of research focused on an existing conceptual core. In this sense, an S-event can deepen the internal activity of a scientific epistemic community by increasing the number of scientists and papers investigating an already established network of concepts. The second process is broadening, which extends scientific inquiry by linking established research to new conceptual networks. Here, an S-event may stimulate other epistemic communities to engage with emerging concepts, thereby expanding the scientific boundaries of a field. The third and potentially most consequential process is interactive, which captures the dynamic interplay between deepening and broadening. In this case, the S-event may catalyze the incorporation of novel concepts into the existing conceptual networks—amplifying the overall connectivity and reach of the epistemic community.

Next, the potential advantages of the S-event approach in observing processes of scientization are outlined. Then three central research questions—one each on deepening, broadening, and a possible dynamic interaction between them—are developed along with operationalization of measures. Finally, to illustrate the empirical application of this approach, the answers to these questions are examined by tracing a multiyear response in scientific papers following an S-event of a transformative research instrument within the education subfield of the broader learning sciences discipline.

Scientific events and scientization

Applying computational analysis to a qualitatively-judged meaningful yet manageable context of a specified S-event—theory challenge, disruptive result, external crisis, or novel instrumentation—offers several advantages over past research on scientization dynamics. Traditional in-depth observations of scientific activity have relied on historical methods that focus on exceptional periods, individuals, or findings (e.g., Mayr, 1982). These studies often highlight revolutionary periods, groundbreaking events, or influential figures, such as the emergence of transformative theories (e.g., the modern synthesis of species evolution and genetics), potentially disruptive hypotheses (e.g., the hierarchy problem in particle physics), or singular scientific geniuses (e.g., Charles Darwin, Isaac Newton, Albert Einstein, Marie Skłodowska Curie). While these subjects are undoubtedly significant, they are often too complex to trace in terms of the routine dynamics of research and its role in furthering scientization (Gauchat, 2024).

An empirical advantage of the S-event approach lies in its ability to preserve the detailed context of scientization while minimizing distortion by avoiding overly exceptional cases. The approach is well-suited for examining more routine yet notable cases—events significant enough to merit attention but not so monumental as to become unwieldy for analysis. In practice, however, the boundaries among different types of S-events can be ambiguous, and these events often interact in dynamic ways, and different types of events may have potential for different responses. Furthermore, every new paper, concept, instrument, or external shock has the potential to qualify as an S-event. Selecting a notable, traceable, and analytically manageable S-event, therefore, requires expert input from active scientists in the relevant research streams. Their expertise is also essential for interpreting the historical flow of papers, citations, and concepts that emerge following a selected S-event.

The advent of digitized platforms containing extensive bibliometric data—such as journal paper titles, authorship, laboratory affiliations, and abstracts—spanning the past 120 years has revitalized computational techniques for analyzing scientific progress. And science-of-science analyses of data from millions of papers have illuminated heretofore unknown sweeping trends in the development of science, but at the cost of mixing thousands of thematic research streams, making it difficult to observe specific processes of scientization (e.g., Sugimoto and Larivière, 2018; Wang and Barabási, 2021). Therefore, a second advantage of the S-event approach lies in applying these methods to the flow of subsequent papers and their semantic networks after a specific event, thus focusing observation on a contained case of routine scientific modus operandi, revealing the micro-dynamics at play in scientization. As the scientific journal article has become the "generalized symbolic medium" for disseminating research (Baker et al. 2025), quantifying attributes of paper lineages—including authorship, content, timing, location, citations, and conceptual connections—provides a reliable and valid measure of a research stream's conceptual coherence, complexity, propensity for innovation, and the timing and pace of its evolution.

PISA as an S-event

In consultation with four active learning scientists, selected here is a notable S-event of innovative instrumentation in the field of education research's subject area of learning science. The 1999 Program for International Student Assessment—PISA as it is commonly known—is an open-access international dataset of psychometrically complex assessments of learning of scholastic mathematics, science, and language skills of ~700,000 15-year-old students from representative samples drawn in 81 countries, with new waves refreshing the dataset every 3 years. In its scale, technical specifications, linguistic demands, and required resources, PISA is significantly beyond any one national collaborative team's capacity. Therefore, international teams of psychometricians, curricular specialists, and learning scientists designed a state-of-the-art, multi-level survey that, along with learning assessments (achievement tests), included information about the students, their schools and families, and the national education system. The design was coordinated and fielded by the multinational Organization for Economic Co-operation and Development (OECD), assisted by sub-contracted survey research firms, national governments, and research institutes. The initial impetus behind PISA was to produce technically sophisticated, internationally comparative reports on the quality of secondary schooling and student outcomes. Over the past quarter-century, PISA's reports have become a fundamental part of educational development and policy analysis for countries, multilateral aid agencies, and international organizations. The frequently influential policy reports are, however, not included here as the focus is on scientization, for which papers are the more appropriate measure (e.g., OECD 2000).

The PISA dataset was made up of thousands of pieces of information coded as quantitative and qualitative variables for statistical modeling employed in new research on a range of education and learning science concepts not necessarily connected to the data's policy mission. In the role as a unique instrument that can facilitate original research, PISA is analogous to other open-access scientific datasets, such as the National Center for Biotechnology Information's GenBank of genetic data.

PISA serves as a compelling case of the S-event approach to observing scientization. Over 18 years since 1999, the learning sciences community has produced over a thousand scientific papers using PISA data, providing a reasonably sized body of research to trace and analyze. Its utility is further enhanced by its unique origins: designed largely independent of specific theoretical schools, so it includes a wide variety of data, plus its scale and complexity were disruptive for their time. Although PISA builds on earlier instruments, it introduced distinct innovations that offer previously unexplored scientific opportunities (e.g., Fishbein, Foy and Yin 2021). Lastly, PISA's state-of-the-art data accessibility is designed to facilitate use by a wide range of researchers, including those new to this type of instrument and to the educational subfield of learning sciences Indeed, in some countries, PISA drove the establishment of a new organizational field of "empirical educational research," including multidisciplinary professional associations (Zapp and Powell 2016).

Operationalizing scientization and research questions

Deepening and broadening of scientific inquiry and their potential interaction can be operationalized to understand the complex scientific responses to a selected S-event. Building from this, three research questions guide the empirical assessment of an S-event approach applied to the PISA case. The S-event approach is organized by bibliometric and then semantic network analyses. Thus, the first and second research question focus on deepening and broadening, and are answered first with bibliometric analyses and then by semantic network analyses. The third research question is on the interaction between deepening and broadening dynamics is answered with semantic network analysis.

Deepening of scientization. Following a significant S-event, scientization deepens within a specific research area with the growth of foundational components of what is often summarized as a scientific epistemic community. The emergence of a new epistemic community or the increasing robustness of an existing one is the key indicator of a deepening response—evidenced by growth in the number of papers, a concentration of papers in specific journal subjects, and an increase in active researchers (Meyer and Molyneux-Hodgson 2010). While there is yet insufficient research on common S-events to establish a standardized metric for evaluating the size and growth rates of these components, a small volume would suggest shallow scientization. Conversely, assuming moderate or greater impact, the rate and pattern of growth also provide a measure of the depth of scientization spurred by the S-event.

Undergirding a growing epistemic community is a network of interrelated research ideas, concepts, terms, substantive phrases—what is referred to here as a network of concepts—contributed by papers that provide an assessment of deepening (concepts can also be analyzed to indicate broadening as described below). Multiple co-occurring concepts and their interconnections indicate an accumulating main core of concepts that are the central knowledge domain of the particular scientific epistemic community. Over time, as more concepts are discussed together, new parts of the knowledge domain can be integrated into the core by researchers bridging formerly less connected research areas. Subfields of research, and perhaps larger main fields, typically begin with a few foundational concepts that expand as new connections are made. The concepts that persist in the main core are those that act as attractors for other concepts: the more interconnected they become, the more likely they are to endure (Cheng et al. 2023).

RQ1. *Did the PISA S-event lead to the emergence and growth of a scientific epistemic community, accumulating a main core of identified research concepts?*

This is answered by analyzing the volume, annual frequency, and growth rates of papers examining aspects of the PISA data over the 18-year period following the instrument's first release of data (1999 to 2017). A similar analysis examines the journals publishing PISA-related papers, both the accumulation and recurrence of papers on PISA. The assessment also considers the growth of the PISA scientific community, as reflected both in the growth in contributions by PISA-active scientists and in the inclusion of new contributors. Thematic and semantic analysis of titles and abstracts from PISA-related papers was verified by experts in education and learning sciences. Using advanced natural language processing and semantic network generation techniques facilitated analysis of PISA's semantic network and yielded evidence about an emerging main core of PISA-related research concepts that indicate deepening scientization in two ways. First is the PISA epistemic community's ability to generate a flow of unique concepts relative to the growth in papers—a precondition. Second is evidence of the accumulation of a main interconnected core of concepts over the period. Main core concepts are operationalized as concepts that are well-connected within the semantic network, appearing in multiple papers and having significant connections (co-occurrences) with other concepts.

Broadening of scientization. Broadening is when subsequent inquiry linked to the S-event progressively explores research concepts at increasing cognitive distance from the main core of concepts. First, this is reflected by the growth in the publishing of S-event-related papers in journals whose subjects are progressively thematically distant from the main subfield. Second, broadening is also indicated by growing numbers of distal—peripheral, isolated—concepts in the semantic network. By themselves, the accumulation of peripheral concepts indicates modest broadening from an S-event, but their volume is also essential for a scientizing interaction between core and peripheral concepts explained next.

RQ2. *Did the PISA S-event lead to, and maintain, significant research peripheral to education learning science and concepts distant from its main core of research?*

This is answered first by expert-assessed bibliometric data on the growth and pattern of PISA papers by journal topics to determine if subsequent research becomes progressively distant from education learning science. Broadening is further indicated if the semantic network, over time, includes research concepts of papers that are peripheral to core concepts of the PISA. Peripheral concepts are operationalized as those that—at a specific time—are isolated in the PISA semantic network: they appear in one or just a few papers without discursive connections to other concepts. Lastly, the most mature broadening process—differentiation—occurs if, over time, sets of peripheral concepts interconnect to form separate cores distal from the main core.

Dynamics between deepening and broadening. A potentially influential mechanism of scientization we test is the dynamic interplay between the deepening and broadening of core and peripheral research concepts within networks of S-event-related papers. When an S-event catalyzes both processes, it can result in a central core of concepts alongside a stream of peripheral concepts, with the potential for interaction between them over time. Once peripheral concepts are introduced in papers, they form a pool from which some may transition to and expand the main core, while others may coalesce into new distinct cores of interrelated concepts distinct from the main core, while still others remain disconnected and peripheral. The former two processes contribute to scientization, but the second is particularly significant as it can drive the emergence of new research concentrations, potentially forming new subfields or redirecting existing ones. Possible interactions between depth and breadth of concepts in a semantic network from the content of papers are illustrated in four heuristic network scenarios in Fig. 1. As shown in concept network A, the content of papers after an S-event can be judged as converging towards a single dominant paradigm if there is a relatively dense core of interconnected concepts and only sparse and disconnected peripheral concepts. Conversely, as shown in concept network B, research may maintain multiple distinct but interconnected areas of focus, or fragment into increasingly specialized and disconnected subfields.

RQ3. To what degree did the PISA S-event lead to a dynamic between deepening and broadening of research?

If the PISA epistemic community develops enough — in the nearly two decade period investigated here — to produce both deepening and broaden of research concepts, an interaction within the network of PISA concepts would be indicated by the semantic network resembling either scenarios B or C more than either A or D.

Methods: tracing the scientific response to an S-event

Bibliometric data. Beginning with the year 1999, immediately after the release and accessibility of PISA, a two-step search process was conducted using the Elsevier Scopus database. First, the keyword "PISA" was used to identify all papers referencing the PISA dataset. Irrelevant results, such as papers unrelated to PISA data (e.g., those about the Italian city of Pisa, the Leaning Tower of Pisa, or Pisa Sporting Club), were excluded, retaining only peer-reviewed journal papers. Second, all remaining papers were reviewed by four learning scientists to confirm that each

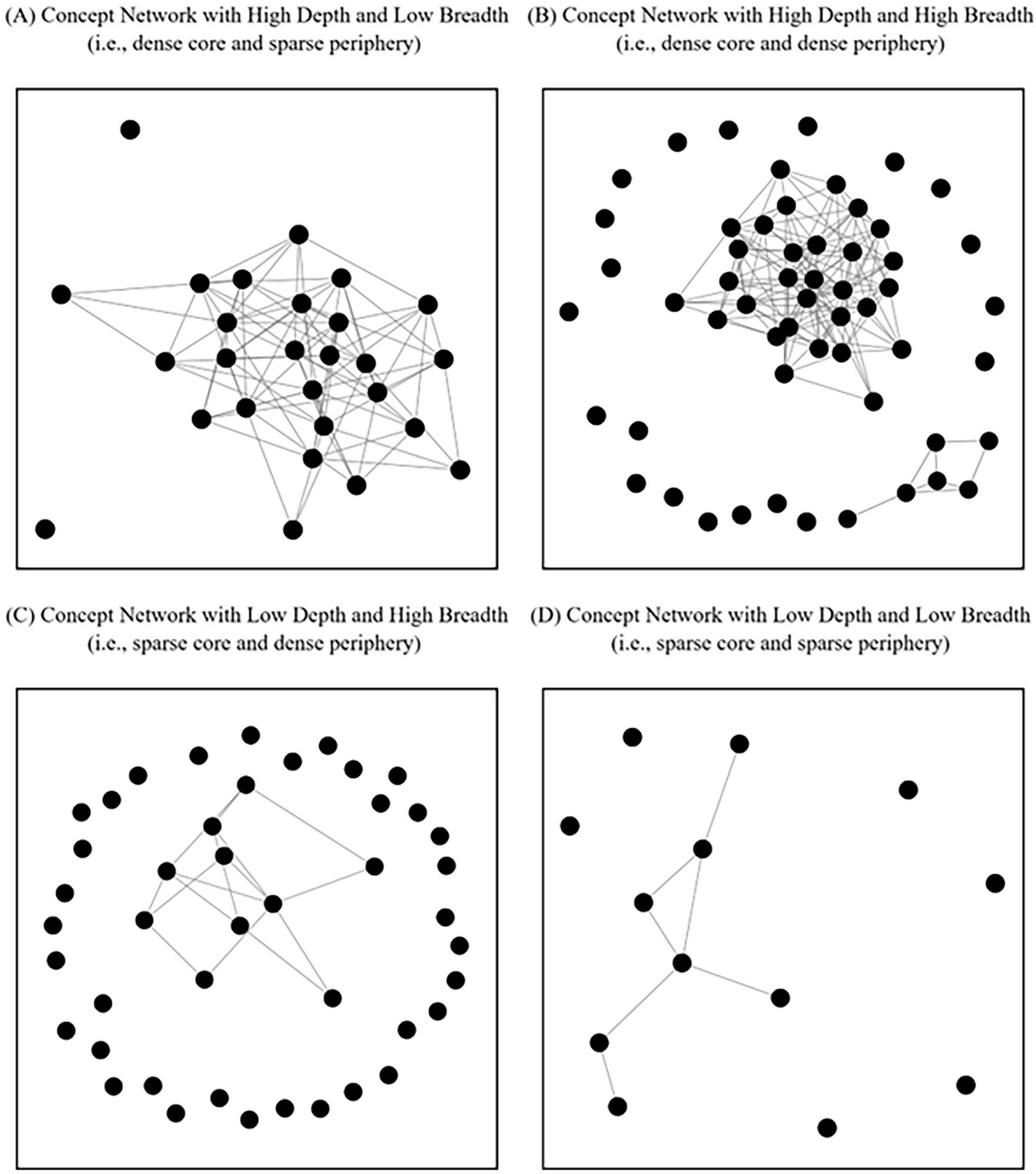

Fig. 1 Heuristic concept network scenarios illustrating how post-S-event research can vary in conceptual depth (density of a central core) and breadth (connectivity/structure of peripheral concepts). Nodes represent concepts and edges represent observed linkages among concepts (e.g., co-occurrence/semantic association within the paper set). **A** High depth, low breadth: a dense, cohesive core with only a few sparse or disconnected peripheral concepts, consistent with convergence toward a single dominant paradigm. **B** High depth, high breadth: a dense core accompanied by a relatively dense/structured periphery, suggesting multiple active research directions around the core and the potential formation of additional interconnected clusters. **C** Low depth, high breadth: a relatively sparse core with many peripheral concepts (largely weakly connected to the core), indicating broad exploration with limited consolidation into a single central framework. **D** Low depth, low breadth: sparse connections both in the core and the periphery, indicating low overall conceptual integration with many isolated concepts.

paper analyzed some aspect of the PISA data, excluding those that merely referenced PISA without substantive analysis. This method identified 1148 papers published from 1999, when the first PISA-related paper appeared, to 2017, a cutoff year chosen to account for citation lag (following Moradel-Vásquez et al., in prep). Metadata for each paper—including title, abstract, authors, author affiliations, journal, and other attributes—was accessed using the *pybliometric* module (Rose and Kitchin 2019). Employing Scopus' All Science Journal Classifications of journals, each journal publishing at least one PISA paper was categorized into three subjects: Education, if the mentioned subject is only Education Learning Science; Education Plus, if the subject category is Education and another major field, such as Education and Psychology; and Others, if the subject category does not include education. Bibliometric analysis of the 1148 PISA papers and their journals evaluates the development of a PISA epistemic community and its deepening and broadening publishing patterns.

Concept extraction and semantic network creation. To analyze the evolution of primary research concepts in PISA papers following the S-event, we developed a computational pipeline integrating transformer-based Natural Language Processing (NLP) with semantic network analysis. We employed KeyBERT, a transformer-based keyword extraction model, to identify representative concepts from the titles and abstracts of each paper. We chose this method because it leverages rich BERT embeddings to identify key phrases that maximize semantic relevance to the full document content. For an academic paper, it means the concepts that the paper tries to cover. This approach offers significant advantages over manual coding (addressing problems of subjectivity and inefficiency) and provides the scalability and consistency necessary for our large-scale analysis.

For implementation, we used the keyphrase-vectorizers library for all text pre-processing (tokenization, stopword removal, lemmatization) and restricted the analysis to English-language papers. We selected a state-of-the-art sentence transformer, *NovaSearch/stella_en_1.5B_v5*, as the embedding model and set the n-gram range to (1, 3) to extract meaningful phrases. Crucially, we employed maximal marginal relevance (MMR) with a diversity value of 0.3. This was essential to balance semantic relevance with diversity, ensuring the selected keywords were collectively representative without being semantically redundant (e.g., avoiding slight variations like "student achievement" and "academic performance" for the same paper).

A key methodological decision was setting the number of extracted concepts to $k = 3$. This choice was not an arbitrary cutoff but the result of an iterative optimization, balancing conceptual resolution with methodological parsimony. As detailed in our sensitivity checks (see Appendix), higher k-values (e.g., $k = 5$, $k = 8$) consistently introduced noise by "overpredicting" concepts. This "overprediction" created redundant links based on peripheral themes, resulting in immediately dense networks that obscured the very dynamics of deepening and broadening we aim to study. Our chosen $k = 3$ configuration, in contrast, provided the most coherent and interpretable map of the field because it directly operationalizes the "conceptual triad" in our theoretical model (see Fig. 2). In this model, we posit that two concepts ground the paper in the existing literature (serving as a "bridge," contributing to deepening), while the third represents its novel contribution or unique focus (serving as an the "extension," contributing to broadening). This aligns with scholarship arguing that innovation often occurs through novel combinations of existing ideas (e.g., Uzzi et al., 2013). This model, in turn, dictates our network linkage threshold: requiring a two-concept overlap (i.e., 66% of a paper's conceptual core) ensures a strong semantic linkage. This high threshold is deliberately chosen to minimize spurious connections that can arise from single, potentially polysemous, keyword overlaps, thereby increasing the robustness of the resulting network structure and ensuring we are mapping meaningful intellectual connections, not fleeting similarities (see Newman, 2018).

We used the extracted concepts to construct a network specifically designed to model our central research goal: to operationalize how a coherent scientific field emerges after an S-event, consolidates over time, and contributes to the deepening and broadening of the scientific domain. We are interested in the mechanism of this consolidation. We theorize that it occurs as individual papers become "knitted together" when their newly generated ideas are successfully integrated into the domain's core conceptual network. This integration, we argue, is achieved through a synthesis with existing, established ideas—a process where the shared existing concepts act as the necessary bridge to pull the new concept into the core.

It is crucial to clarify that our use of "core" and "periphery"—terms central to this bridging process—differs explicitly from the traditional, structural definitions common in network science (cf. Zinilli 2025). Standard core-periphery algorithms identify a dense, tightly-knit core and a sparse periphery within a static network snapshot. Our approach, however, is fundamentally dynamic and process-oriented. A concept enters the "core" in our model, not based on its final position in a static structure, but through the cumulative process of being used to bridge papers over time through their shared conceptual territory. The "core" simply represents the connected nodes that have been integrated into the network, while the "periphery" represents the pool of concepts that have not yet been integrated via this bridging mechanism. This process-based definition is essential for our research question, which is about tracing the evolution of knowledge through deepening (a growing network "core") and broadening (the integration of "peripheral" concepts). Furthermore, while standard concept-concept networks are useful for showing which concepts are related and concept-paper bipartite networks can map which papers use which concepts, neither of these models directly captures this dynamic process of consolidation. A concept-concept network, for instance, provides a static map of the conceptual landscape but abstracts away the papers doing the work, failing to show how those papers build a shared "semantic core". A bipartite network is similarly insufficient; while it would show links from "Paper 1" to "Concept A" and from "Paper 2" to "Concept A", the crucial relationship between "Paper 1" and "Paper 2" remains merely a derived, two-step path (Paper 1 → Concept A → Paper 2) rather than a direct object of analysis. This structure shows similarity, but it fails to model the agglomerative process of these two papers merging to form a new, single semantic entity.

Considering these elaborations, we operationalized a paper-paper network where the nodes are the papers themselves. In this model, the concepts are not the nodes but rather the explicit mechanism for creating links. This approach provides the critical advantage of allowing us to directly observe the agglomeration of the field. We can watch how a new paper (a new node)—acting as a vehicle for new ideas—integrates into the main body of knowledge precisely because its overlapping concepts (its shared "conceptual DNA") function as the bridging mechanism that connects its new ideas to the existing core. This is what allows us to test our theory about "long-term survival and consolidation" (Cheng et al. 2023)—it is not just about concepts co-occurring, but about the progressive integration of papers via these conceptual bridges. This methodology, grounded in the principle that co-occurrence signifies conceptual alignment (Harris, 1954; Jurafsky and Martin, 2009; Gärdenfors, 2014), is described in detail below.

As depicted in Fig. 2, the construction of the semantic network operationalizes our core theory of field consolidation. The process begins with individual papers (P1, P2, P3) as isolated subgraphs, each containing its three extracted concepts (e.g., P1: {A, B, C}; P2: {A, B, D}; P3: {C, D, E}). The process of establishing inter-paper connections hinges on our key threshold criterion: a tie is formed, and subgraphs are merged, if at least two shared concepts are identified. This $k = 3$/overlap $= 2$ configuration is the crucial mechanism, as it models our hypothesis: It ensures a new idea (the third, unique concept) can only be integrated if it is bridged by a sufficient foundation of existing, shared concepts (the two overlapping ones). For example, P1 {A, B, C} and P2 {A, B, D} share the "existing" concepts "A" and "B". This satisfies the threshold, and their subgraphs merge into a single semantic core {A, B, C, D}, successfully integrating the "new" ideas "C" and "D" into one component. In a subsequent step, this new core {A, B, C, D} is compared to P3 {C, D, E}. They share concepts "C" and "D"—which now function as the established bridge—satisfying the threshold. This allows P3 to integrate, successfully bringing its new concept "E" into the fully connected semantic network {A, B, C, D, E}.

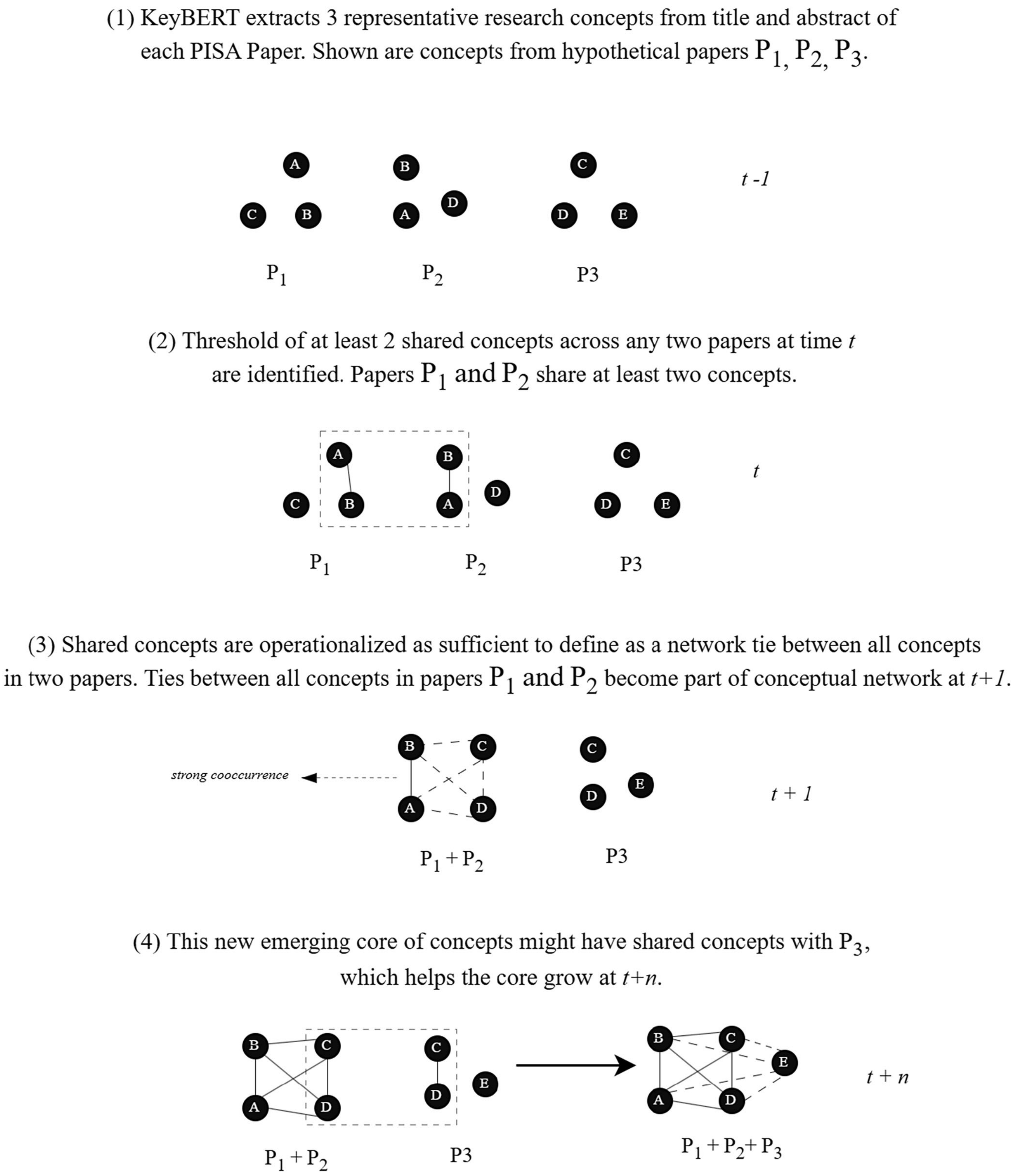

Fig. 2 Schematic of how paper-level concept extraction and overlap rules generate concept network ties over time. Each paper contributes $k=3$ KeyBERT concepts (two "bridge" concepts anchoring prior work and one "extension" concept capturing novelty). (1) Concepts are extracted from each paper (illustrated for hypothetical papers P1–P3). (2) Paper pairs are linked only when they share at least two of three concepts (a 66% overlap), identifying strong semantic similarity at time *t*. (3) When a qualifying overlap occurs (e.g., P1 and P2), ties are added among all concepts in those papers, forming an emerging core at $t + 1$. (4) Additional papers that later share concepts with the core (e.g., P3) can be incorporated over time ($t+n$), expanding the network.

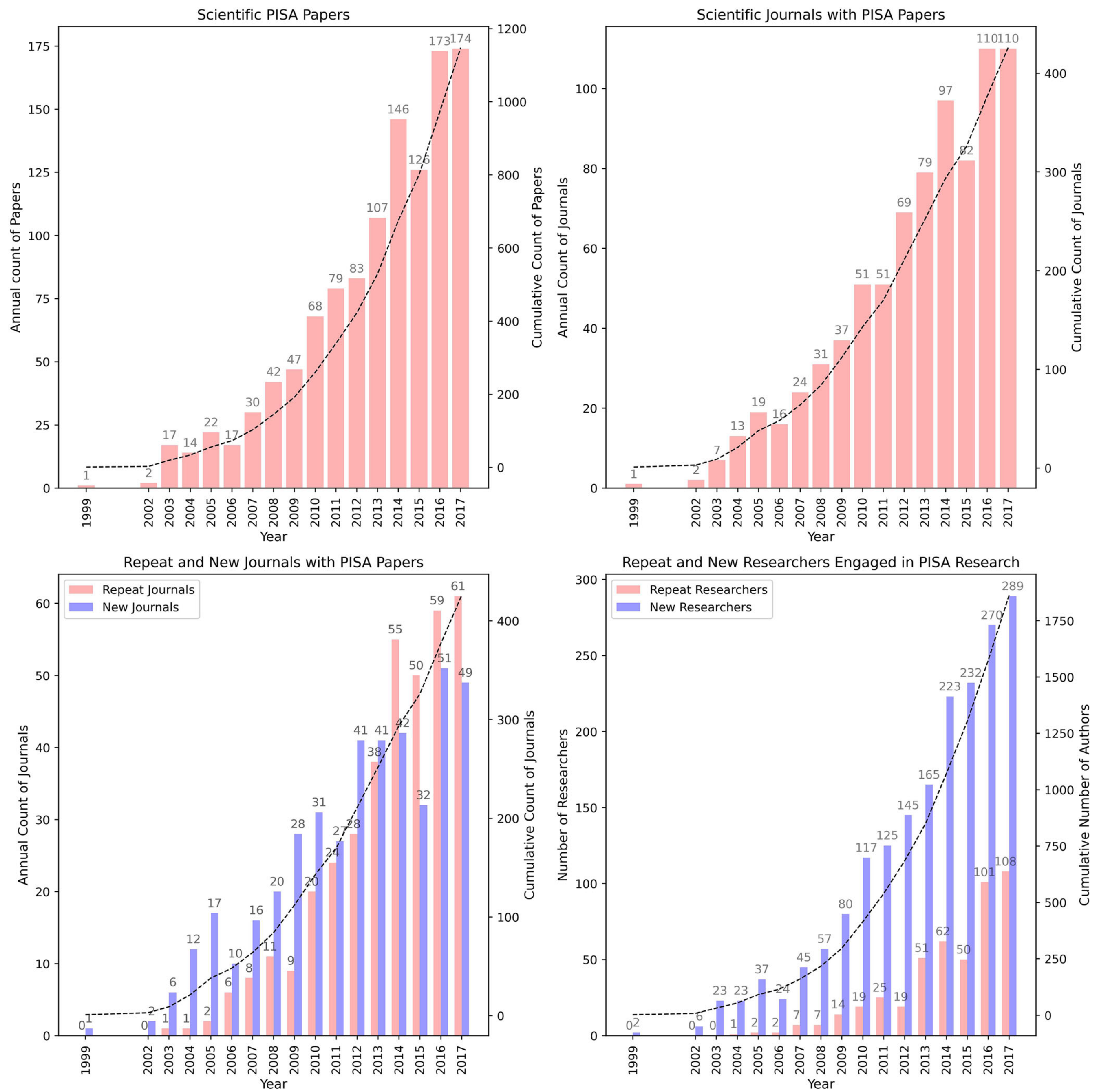

Fig. 3 Growth of the scientific literature on PISA, showing annual counts (bars) and cumulative totals (dash lines) from 1999–2017. A Annual and cumulative number of PISA papers. **B** Annual and cumulative number of journals publishing PISA papers. **C** Annual counts of repeat journals (those that have previously published PISA papers) versus new journals, with cumulative total journals overlaid. **D** Annual counts of repeat versus new researchers publishing on PISA, with cumulative authors overlaid.

The iterative nature of this approach models the progressive growth of thematic cores, capturing both immediate conceptual relationships within individual papers and broader structural integration of ideas across the corpus over time. By transitioning to a concept-based framework, this methodology enables a more granular understanding of thematic depth and breadth, visualizing how research themes emerge, evolve, and interconnect within the global PISA landscape.

Expert scientists' assessments. Four learning scientists, three of whom are knowledgeable about PISA dataand one who was involved in its design and has authored several scientific papers using the data, contributed their expertise to three aspects of the explained S-event approach:

1. Verification of paper inclusion: The experts confirmed whether the identified papers qualified as PISA research papers by reviewing their titles, abstracts, keywords, and, when necessary, their full text.
2. Validation of Elsevier Scopus journal classification: To assess the classification of journal topics relative to PISA, the experts evaluated a randomly selected sample of papers from journals outside the core "education and learning sciences" category. They verified that PISA data was actively analyzed (not merely mentioned) and that the analysis was

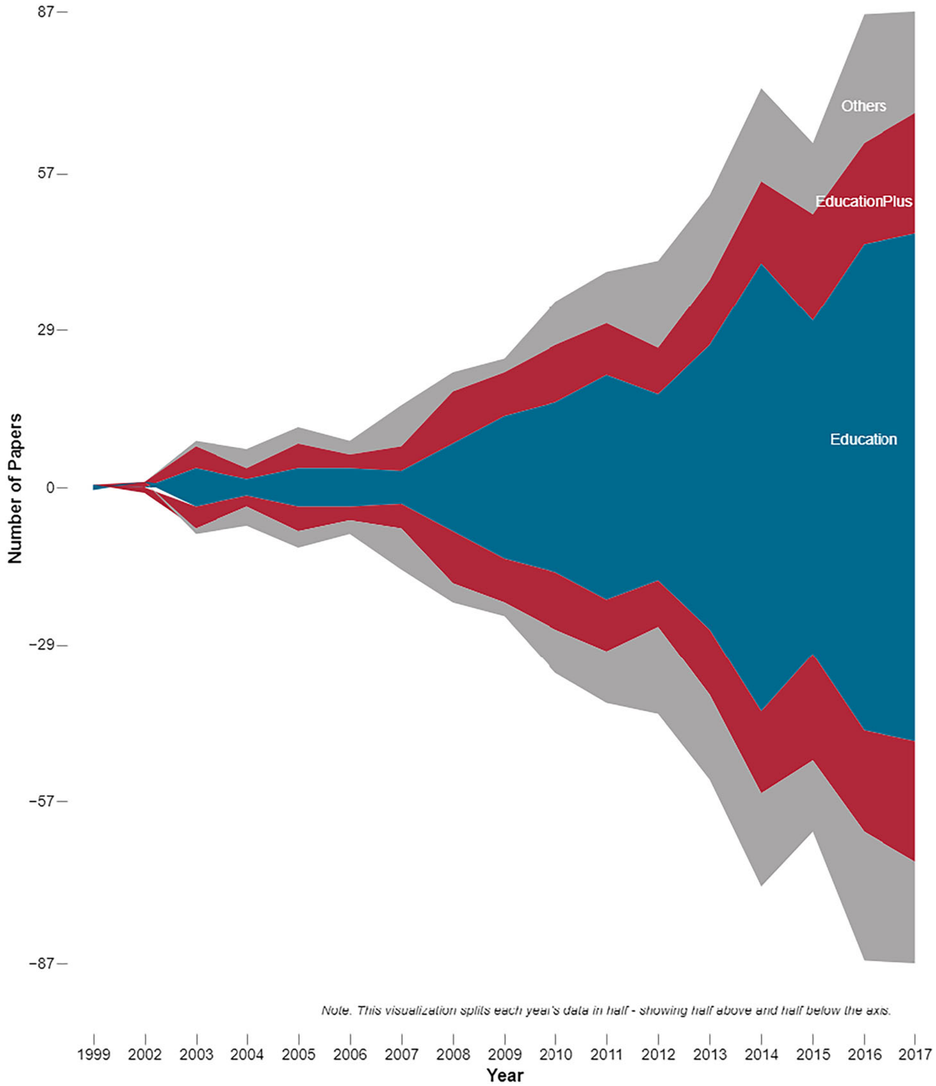

Fig. 4 Disciplinary composition of PISA-related papers over time (1999–2017), showing how annual paper counts are distributed across Education, Education Plus, and other fields. Colored bands represent the number of papers in each category per year; the display splits each year's total so that half is plotted above and half below the horizontal axis.

applied to concepts beyond education and learning sciences. All sampled papers met these criteria.

3. Validation of semantic network methods: The experts reviewed selected KeyBERT-identified main core and peripheral concepts to assess the external validity of the method for determining the semantic network. There was high agreement on concept classification between the experts and the computational method.

Results

Using the PISA case to illustrate the S-event approach is organized by bibliometric and then semantic network analyses. The first and second research questions on deepening and broadening, respectively, are answered first with bibliometric analyses and then by semantic network analyses. The third research on the interaction between deepening and broadening is answered with just semantic network analysis.

Bibliometric results. The PISA S-event in 1999 resulted in the formation of an active epistemic community. As displayed in the upper left-hand graph of Fig. 3, PISA facilitated a stream of scientific papers with the annual frequency following a maturing logistic growth curve with slow growth of 3.6% annual increase in papers until an inflection point at 2006 and an accelerating rate of 27% annually until 2014 followed by a slowing to 1% annually approaching an asymptote of research with the PISA data. This resulted in an accumulating annual growth rate of ~11–12% from 2003 to 2006, increasing to 13–16% from 2007 to 2009, and reaching over 30% annually from 2015 to 2017. A similar logistic pattern is evident in the upper-right-hand figure, which displays the annual growth in journals publishing PISA papers, culminating in 430 unique journals over the period. This growth is accompanied by a wide dispersion of papers, as reflected in a relatively modest ratio of just over two and a half papers (2.7) per journal. The lower-left-hand graph, however, indicates the gradual emergence of a core set of journals; early in the scientific response, papers in journals that had published at least one prior PISA paper dominated annually over new journals up to 2013, after which papers in journals new to PISA became the modal pattern. Because there are approximately 700 journals dedicated to research on education learning sciences and acceptance rates are low among many, counting only those journals publishing at least two PISA paper may underestimate the development of a

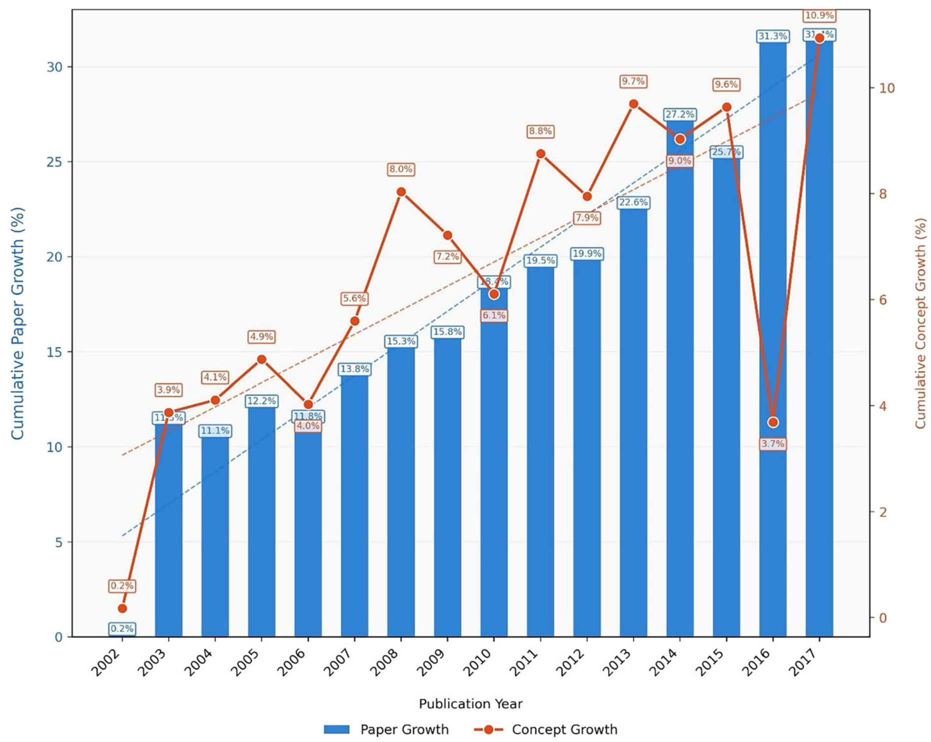

Fig. 5 Cumulative weighted annual growth rates of PISA papers (bars) and new concepts (line), 2002–2017 (base year 1999). Paper growth increases steadily over time, while concept growth fluctuates, including a marked dip in 2016 (fewer new concepts per paper).

core of journals. The lower-right-hand graph of Fig. 3 indicates a slowly forming core of scientists until 2013, when 51 papers that year included at least one author who had already authored a prior PISA paper, rising to just over 100 such papers by the end of the period. As also shown, PISA attracted a growing and significant number of new scientists annually, particularly from 2014, but with a modest number of repeat authors.

An emerging and growing core of journals becomes visible when we examine every education-related journal with at least one PISA paper. Figure 4 presents the Scopus and expert-verified journal classification of all PISA papers depicted as a river graph of journals publishing PISA papers. Journal classification appears to fall into one of three widening streams representing increasing cognitive distance in the use of the PISA data. Over time, the epistemic community published in journals with topics solely on learning sciences and in journals on topics progressively distant from purely educational. To operationalize this, we collapsed Scopus/ASJC assignments into three overarching groups as follows: (i) “Education” includes records whose Scopus subject-category assignment is exactly “Education” (ASJC: Education; code 33043304, as labeled in Scopus); (ii) “EducationPlus” includes records where “Education” appears among multiple Scopus subject-category assignments (i.e., Education co-assigned with at least one additional ASJC category); and (iii) “Others” includes records without an “Education” subject-category assignment. The large central blue stream is growth in papers appearing in journals focused on education. These include: *Revista de Educación, European Educational Research Journal, Comparative Education Review, Zeitschrift für Erziehungswissenschaft, Educational Research and Evaluation, International Journal of Science Education, European Journal of Education, Profesorado, RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, and *Journal of Education Policy*. PISA papers in these journals tend to focus on analyses of the associations between various educational inputs — national spending on education or distributions of teacher qualifications—and outputs, usually the achievement tests. The red stream, separated into two for visual effect, is an accumulation of papers in Education Plus journals. In this case, including papers with the PISA data applied to combinations of education and content from other fields; a modest type of breadth increasing from six such journals in 2005 to 36 in 2017. Education Plus includes: *Zeitschrift für Pädagogik, International Journal of Science and Mathematics Education, Economics of Education Review, Intelligence, International Journal of Educational Development, Learning and Individual Differences, ZDM - International Journal on Mathematics Education, Education Economics, Social Indicators Research, Eurasia Journal of Mathematics, Science and Technology Education*. The greatest journal topical breadth is the gray stream where PISA papers are in journals with a primary topical focus other than education. Increasing from six in 2005 to 41 in 2016, these include: *Sociologický Časopis, Applied Economics, Regional and Sectoral Economic Studies, European Sociological Review, Kölner Zeitschrift für Soziologie und Sozialpsychologie, Empirical Economics, Revue Française de Sociologie, OECD Journal: Economic Studies, Frontiers in Psychology, Emotion Review*.

Semantic network results. Analyses of semantic networks—comprising concepts and their interconnections found in paper titles and abstracts—support and enrich the bibliometric findings. An S-event can generate a pool of research concepts that reflect the content of related papers, becoming the essential ingredient for both deepening and broadening. As noted in Fig. 2, when novel concepts are shared enough across papers, they reflect a consistent set of core concepts that can grow—deepen—with subsequent papers by the epistemic community. Less shared new concepts reflect a distal periphery that potentially could develop with subsequent papers—broaden. This creation of a concept pool is shown in Fig. 5, displaying the parallel increase in the

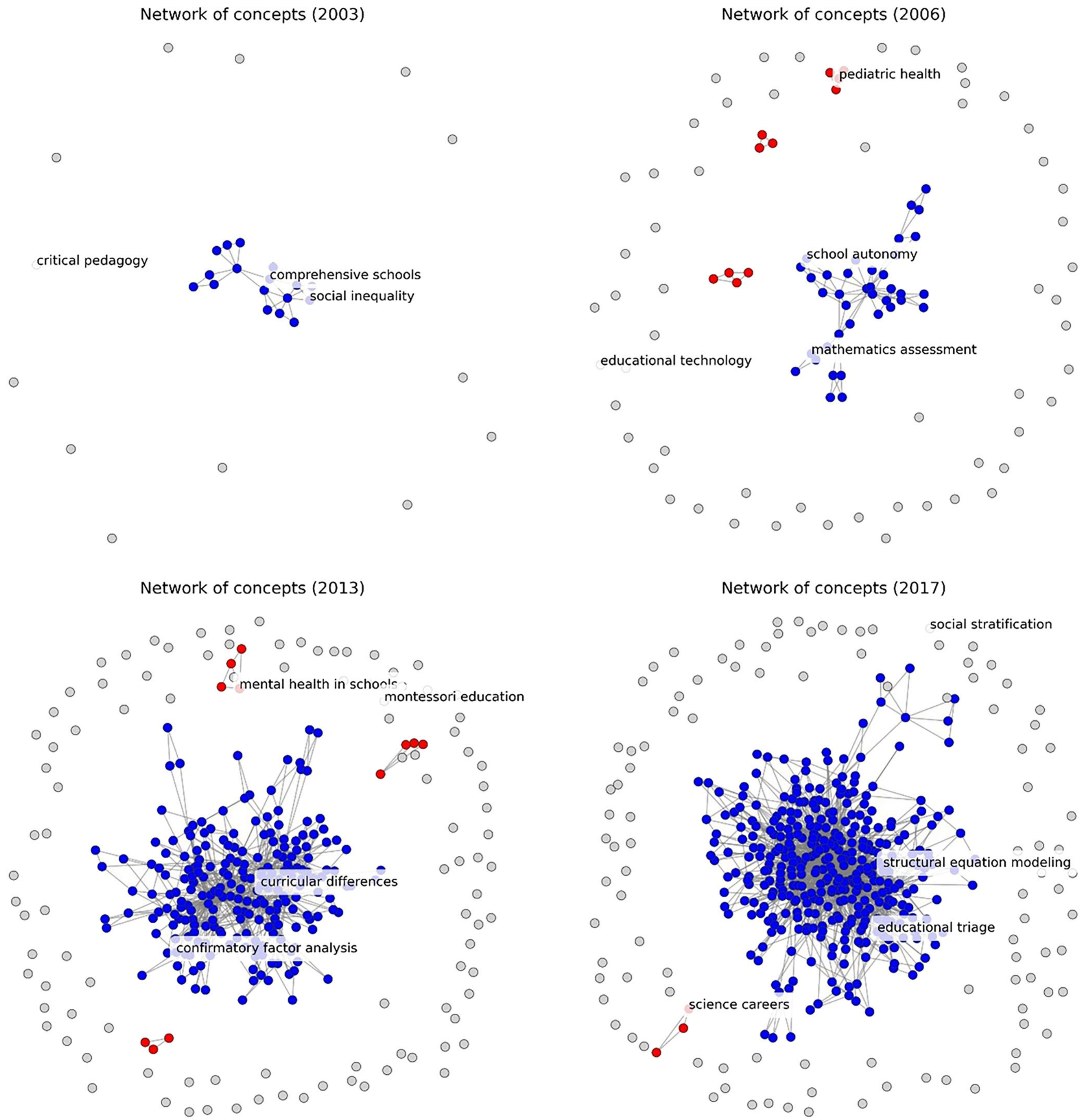

Fig. 6 Semantic networks of concepts in PISA papers at four time points (2003, 2006, 2013, 2017), illustrating core growth and peripheral turnover over the 2002–2017 period. Blue nodes indicate the main interconnected core, gray nodes indicate peripheral concepts, and red nodes indicate small distal (competing) cores formed by concepts shared outside the main core; labels are shown for illustration.

cumulative annual growth rates of PISA papers and new concepts from 2002 to 2017. With an average of ~1.5 unique concepts per paper, the annual growth rate of new concepts into a pool for future research over the whole period was calculated to be approximately 19.65% per year.

To statistically validate these descriptive trends, we performed a linear regression analysis between time and growth in papers and concepts (see Appendix Table 2). The results confirm that the linear trends for both cumulative paper growth ($B = 1.688$, $p < 0.001$) and new concept growth ($B = 0.437$, $p < 0.01$) are highly significant, demonstrating that the S-event triggered a sustained, non-random accumulation of both papers and novel ideas. Furthermore, the R-squared values reveal a distinction in their underlying dynamics. Paper growth is extremely stable and predictable ($R^2 = 0.928$), reflecting a steady, institutionalizing process of conceptual deepening as the epistemic community matured. In contrast, new concept growth is significantly more variable ($R^2 = 0.534$). This lower R-squared value provides statistical support for our argument that conceptual broadening is a less predictable process of innovation, occurring in bursts rather than as a smooth, linear progression. Immediately following the PISA S-event, the growth in papers outpaced the

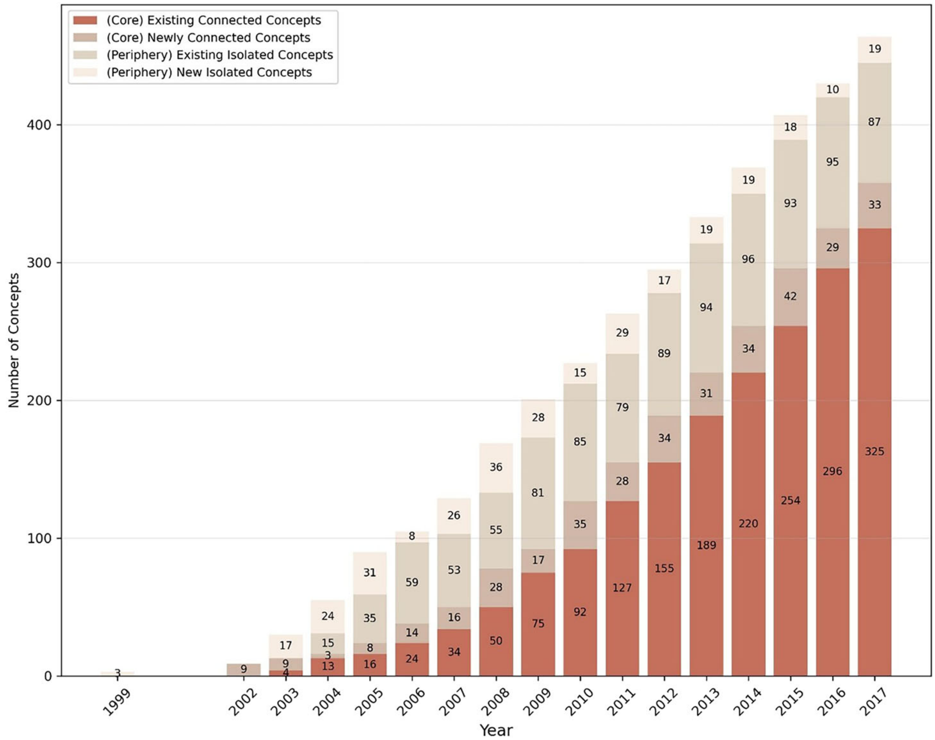

Fig. 7 Annual changes in the PISA concept network showing how the main core deepens and the periphery broadens as new papers add concepts, 1999–2017. Stacked bars partition concepts into: existing core concepts that remain connected, newly connected concepts that join the core, existing isolated (peripheral) concepts, and newly introduced isolated concepts.

introduction of new concepts, a pattern that might be expected as the emerging epistemic community focused on a few shared concepts. This changed as the cumulative growth rate of new concepts introduced in PISA papers rose from 4 to 5% in 2003–2006 to 6 to 8% in 2008–2010 and surged to 9 to 11% in the final four years of the period. The remainder of the analysis of the semantic network focuses on this pool.

As a facilitating instrument, the PISA S-event often introduces new concepts that reflect applications of its data to various areas of education research, which then become part of the conceptual core. For example, across the papers published in 2004 and 2005, the first period of growth in the accumulated rate, new concepts that with more papers eventually entered the conceptual core include school autonomy, mathematic assessment, teacher diversity, functional literacy, adult literacy, nonparametric methods, immigrant integration, cultural capital, human development index, whole child approach, and multicultural education. From 2008 to 2010, the most intense period of growth, new shared concepts developed the core, including excellence clusters, kindergarten readiness, situated learning, intergenerational mobility, school dropouts, and cultural capital. Examples of concepts that remained peripheral from the core include self-actualization, adaptive instruction, parent-teacher associations, and entrepreneurship education. Other concepts, like pediatric health and adult literacy, that were initially in competing cores eventually merged with the main core.

Figure 6 shows the development of the semantic network among concepts as they are introduced and shared in papers by the epistemic community. The most shared concepts formed a small conceptual core by 2003 with a sparse set of peripheral concepts. Two years later, the network had changed significantly with a deepening core of concepts, more peripheral concepts, and evidence of enough sharing among some peripheral concepts to form possible competing distal cores. For example, although the peripheral concept of "pediatric health" was not shared enough with subsequent papers on core concepts, it was shared by a small set of two papers in 2006. The network in 2013 showed the same earlier pattern. Two years later, though, the semantic network had evolved a large shared conceptual core and many peripheral concepts. Also, as later papers in conjunction with earlier ones expand the main core of concepts, competing distal cores were likely to have been subsumed into the main core. For example, the initially distal concept of "Montessori education" was by 2017 sufficiently shared with subsequent papers that also incorporated the main core concepts to make it part of the main core. See Appendix Table 1 for dimensions of the semantic network, including competing distal cores.

Figure 7 shows the dynamics between the conceptual core and periphery within the evolving semantic network. For example, all the PISA papers published through 2012 resulted in an accumulating main core of 189 interconnected concepts and 106 peripheral concepts; in 2013, these increased to 216 and 117 concepts, respectively. This happens from a three-part dynamic within the semantic network stimulated by the addition of concepts from 107 newly published 2013 PISA papers. First, the main core deepened through the inclusion of new concepts that some 2013 papers shared with those already in the core. Second, with this expansion, some formerly peripheral concepts now had sufficient sharing across new papers to be part of the main core. By a combination of these two dynamics, the main PISA conceptual core grew by 27 concepts in this year. Third, the periphery of PISA research broadening by an additional 11 new concepts found in the 2013 papers, and there were three competing distal cores. As noted above, in the case of the PISA S-event, with new papers adding in subsequent years, the first dynamic eventually pulled enough competing distal core concepts into the main core so that by 2017 only one remained (See Appendix Table 1). To illustrate these dynamics, concepts such as curriculum reform, achievement gap, mathematics assessment, gender differences, and competencybased education were in the core and the focus of research across papers early in the period (1999–2005), while concepts like inclusive education (appeared in 2003), critical pedagogy (appeared in 2003), developmental education (appeared in 2004) were in the periphery. They were included in subsequent papers enough to become part of the core by 2007, 2014, and 2017, respectively. Others—ability grouping, social selectivity, and Pygmalion effect—remained in the periphery since their introduction in 2005, 2007, and 2012, respectively, waiting to be shared by the future epistemic community.

Conclusion: Assessing the S-event approach to observing scientization

As expected with the introduction of a large, complex, open-access dataset like PISA, the bibliometric data reveal the formation of a modestly intensifying epistemic community comprising scientists, papers, and journals. The evidence suggests this community has evolved with permeable boundaries, rather than a classic, tightly defined one. PISA papers have been spread across many journals, even within the expected areas of education, learning sciences and the larger field of education. Also, there was a small, growing stream of papers in journals with increasing cognitive distance from education. Across the period, papers in a growing set of journals with repeated PISA contributions were about equal to papers in journals new to publish PISA analyses. The analysis also evidenced the attraction of the PISA data, which brought a consistent influx of new researchers, who significantly outnumbered those with prior PISA papers throughout the period. Despite the open-ended nature of PISA's epistemic community, its fundamental components steadily accumulated over nearly two decades, with growth rates slowing by 2017. The concepts that this fluid PISA epistemic community researched reflect both the deepening and broadening scientific response and dynamic interaction between these processes.

Analyzing the semantic network reveals a core set of concepts that expanded and grew increasingly interconnected in later papers—indicative of a deepening scientific response. Concurrently, the papers introduced a steady stream of peripheral concepts, with some evidence of small, alternative cores forming. These cores, while short-lived in this case, represent potential precursors to the creation of new subfields or significant shifts in existing ones, demonstrating a broadening response. Peripheral concepts from earlier papers frequently moved into the core of later papers, resulting in an accumulated semantic network by 2017, characterized by both significant depth and breadth of research topics. This outcome is likely a result of the open-access nature of the PISA dataset, a characteristic feature of this S-event, coupled with the presence of an open epistemic community with fluid boundaries, also given the relatively low barriers in thematic and methodological terms.

The S-event approach offers a distinct analytical perspective on science, complementing both qualitative case studies of discovery and large-scale data analyses across extensive scientific domains. While additional S-event cases are needed to compare the magnitude of scientific responses, the PISA case provides feasible and interpretable evidence regarding the process of scientization. The advent of big data and computational capacity has strengthened entry into the epistemic community of PISA research. It is also unsurprising that the semantic network in this specific case included many new concepts over time across papers. Also, this illustration employed a liberal sharing criteria of just two concepts across two papers, making for an inclusive conceptual core. Larger and hence more conservative sharing criteria can be selected and compared depending on specific hypotheses about other types of S-events. Future studies could replicate these findings with larger datasets and use the structural analysis prevalent in network science to validate these analysis at the macro-level (Zinilli 2025). Also, it is important to note that this journal article-focused analysis captures only a portion of PISA's broader impact as an S-event, excluding OECD technical reports, policy briefs, national evaluation documents, and other non-indexed papers that substantially contribute to the PISA knowledge ecosystem. Future research could productively expand to analyze this broader publication landscape.

Tracing scientific response to an S-event proved practical, particularly with the support of scientific experts to validate the inclusion of papers, journal topics, and concepts. By combining bibliometric analysis with semantic networks of concepts, the approach yielded novel insights into the formation of an epistemic community and its contributions to both deepening and broadening scientific knowledge. Most notably, this case highlights potential dynamics between these two forces, offering a new perspective on their role in advancing scientization.

Data availability

The raw data for this study were collected from the Scopus database, available via the Scopus API. The complete replication package, including the curated dataset and analysis code, is openly available in the Harvard Dataverse repository at: [https://doi.org/10.7910/DVN/NMZQUX].

Received: 23 April 2025; Accepted: 2 January 2026;
Published online: 23 February 2026

References

Adler E, Haas P (1992) Conclusion: epistemic communities, world order, and the creation of a reflective research program. Int Organ 46:367–390

Baker DP, Powell JJW (2024) Global mega-science: Universities, research collaborations, and knowledge production. Stanford University Press

Baker DP, Powell JJW, Adeel AB, Marques M (2025) Is more science really less science? The scientization of science, 1900–2020. Minerva https://doi.org/10.1007/s11024-025-09602-1

Bloom N, Jones CI, Van Reenen J, Webb M (2020) Are Ideas Getting Harder to Find? American Economic Review 110(4):1104–1144

Bornmann L, Haunschild R, Mutz R (2021) Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases. Humanit Soc Sci Commun 8:224

Cheng M, Smith DS, Ren X, Cao H, Smith S, McFarland D A (2023) How new ideas diffuse in science. Am socio rev 88(3):522–561

Collins H (2018) Gravity's kiss: the detection of gravitational waves. MIT Press

Dong Y, Ma H, Shen Z, Wang K (2017) A Century of Science: Globalization of Scientific Collaborations, Citations, and Innovations Association for Computing Machinery, https://doi.org/10.1145/3097983.3098016

Drori GS, John WM, Francisco OR, Evan S (2003) Science in the Modern World Polity: Institutionalization and Globalization Stanford University Press

Landhuis E (2016) Scientific literature: Information overload. Nature 535(7612):457–458

Fishbein B, Foy P, Yin L (2021) TIMSS 2019 user guide for the international database 2nd edn. Boston College, TIMSS & PIRLS International Study Center

Fuchs S (1992) A sociological theory of scientific change. Soc F 71:933

Fujita K, Kajikawa Y, Mori J, Sakata I (2014) Detecting research fronts using different types of weighted citation networks. J Eng Technol Manag 32:129–146

Gärdenfors P (2014) The geometry of meaning: Semantics based on conceptual spaces. MIT Press

Gauchat G (2023) The legitimacy of science. Annu Rev Sociol 49:263–279

Gauchat G (2024) Review of Global mega-science: universities, research collaborations, and knowledge production. Soc F. https://doi.org/10.1093/sf/soae138

Glänzel W, Schubert A (2004) Analysing scientific networks through co-authorship. In: Handbook of quantitative science and technology research: the use of publication and patent statistics in studies of S&T systems. Springer, Netherlands, pp 257–276

Haas P (2015) Epistemic communities, constructivism, and international environmental politics. Routledge

Hanson MA, Gómez, Barreiro P, Crosetto P, Brockington D (2024) The strain on scientific publishing. Quant Sci Stud 5(4):823–843
Harris Z (1954) Distributional structure. Word 10:146–162
Holst V, Algaba A, Tori F, Wenmackers S, Ginis V (2024) Dataset Artefacts are the Hidden Drivers of the Declining Disruptiveness in Science. arXiv 2402.14583
Jurafsky D, Martin JH (2009) Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition. 2nd edn. Prentice Hall
Krauss A (2024) Science of science: Understanding the foundations and limits of science from an interdisciplinary perspective. Oxford University Press
Marques M, Macasaet B, Powell JJW, Dusdal J, Baker DP (2025) Scientizing the world: on mechanisms and outcomes of the institutionalization of science. Science Public Policy. https://doi.org/10.1093/scipol/scae095
Mayr E (1982) The growth of biological thought: Diversity, evolution, and inheritance. Harvard University Press
Merton RK (1996) On social structure and science. University of Chicago Press
Meyer M, Molyneux-Hodgson S (2010) Introduction: the dynamics of epistemic communities. Sociol Res Online 15:109–115
Michaelson AG (1993) The development of a scientific specialty as diffusion through social relations: the case of role analysis. Soc Netw 15:217–236
Miller HT, Fox CJ (2001) The epistemic community. Adm Soc 32:668–685
Milojević S (2015) Quantifying the cognitive extent of science. J Informetr 9(4):962–973. https://doi.org/10.1016/j.joi.2015.10.005
Moradel-Vásquez JJ, Macasaet BT, Dusdal J, Pang J, Powell JJW, Fu YC, Adeel AB, Prieto J, Baker D (in prep) Disciplinary Reach of PISA: A Network Analysis of Scientific Knowledge Diffusion. SEEDS Working Paper
Morris SA, Van der Veer Martens B (2008) Mapping research specialties. Annu Rev Inf Sci Technol 42:213–295
Mark N (2018) Networks. Oxford University Press
Organization for Economic Co-operation and Development (OECD) (2000) Measuring student knowledge and skills: the PISA 2000 assessment of reading, mathematical and scientific literacy. OECD. https://doi.org/10.1787/9789264181564-en
Park M, Leahey E, Funk RJ (2023) Papers and patents are becoming less disruptive over time. Nature 613:138–144
Powell JJW, Baker DP, Fernandez F (Eds) (2017) The Century of Science: The Global Triumph of the Research University. Emerald Publishing Ltd
Rose ME, Kitchin JR (2019) pybliometrics: scriptable bibliometrics using a Python interface to Scopus. SoftwareX 10:100263
Schofer E (2003) The global institutionalization of geological science, 1800 to 1990. Am Sociol Rev 68:730–759
Schofer E, Hironaka A (2005) The effects of world society on environmental protection outcomes. Soc F 84:25–47
Sugimoto CR, Larivière V (2018) Measuring research. Oxford University Press
Brian U, Mukherjee S, Stringer M, Jones B (2013) Atypical Combinations and Scientific Impact. Science 342(6157):468–472. https://doi.org/10.1126/science.1240474
Wang D, Barabási A-L (2021) The science of science. Cambridge University Press
Web of Science Group (2022) Web of science core collection. https://mjl.clarivate.com/home
Zapp M (2022) Revisiting the global knowledge economy: The worldwide expansion of research and development personnel, 1980–2015. Minerva 60:181–208
Zapp M, Powell JJW (2016) How to construct an organizational field: Empirical Educational Research in Germany, 1995–2015. Eur Educ Res J 15(5):537–557
Zinilli A (2025) Elements of network science: theory, methods and applications in Stata, R and Python. Routledge

Acknowledgements
The authors acknowledge financial support of the Institute for Advanced Studies of the University of Luxembourg through a BRAINSTORM grant (Brainstorm-01-2022) for "Scientization Events and Expansion across Disciplines in Society" (SEEDS; https://www.uni.lu/research-en/ias/funded-projects/brainstorm/).

Author contributions
All authors conceptualized the paper together. JJMV provided the data. DPB and ABA planned the analysis, analyzed the data, and wrote the manuscript. BTM, JJWP, JP, JD, J.Pang, and YCF revised and polished the manuscript.

Competing interests
The authors declare no competing interests.

Ethical approval
This study did not involve human participants or animal subjects. The research relied exclusively on secondary data retrieved from the publicly available Scopus database via API. As the study involved the analysis of existing, public bibliographic records and did not include any direct intervention or interaction with human subjects, ethical approval from an Institutional Review Board (IRB) or ethics committee was not required according to standard research regulations.

Informed consent
This study did not involve human participants; therefore, informed consent was not applicable. The data utilized in this research consists of bibliographic records obtained from the Scopus database, which are publicly available for research purposes. No private or personally identifiable information requiring individual consent was collected or analyzed.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-026-06490-y.

Correspondence and requests for materials should be addressed to Abdul Basit Adeel.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.