OPEN ACCESS
Advances in Complex Systems
Vol. 26, No. 1 (2023) 2350003 (33 pages)

DOI: 10.1142/S0219525923500030

World Scientific
www.worldscientific.com

BIOMEDICAL CONVERGENCE FACILITATED BY THE EMERGENCE OF TECHNOLOGICAL AND INFORMATIC CAPABILITIES

DONG YANG

Department of Management of Complex Systems,
Ernest and Julio Gallo Management Program,
School of Engineering, University of California,
Merced, California 95343, USA

IOANNIS PAVLIDIS

Computational Physiology Laboratory,
Department of Computer Science,
University of Houston,
Houston, Texas 77204, USA

ALEXANDER MICHAEL PETERSEN*

Department of Management of Complex Systems,
Ernest and Julio Gallo Management Program,
School of Engineering, University of California,
Merced, California 95343, USA
apetersen3@ucmerced.edu

Received 10 February 2023
Revised 27 April 2023
Accepted 10 May 2023
Published 26 June 2023

We leverage the knowledge network representation of the Medical Subject Heading (MeSH) ontology to infer conceptual distances between roughly 30,000 distinct MeSH keywords — each being prescribed to particular knowledge domains — in order to quantify the origins of cross-domain biomedical convergence. Analysis of MeSH co-occurrence networks based upon 21.6 million research articles indexed by PubMed identifies three robust knowledge clusters: micro-level biological entities and structures; meso-level representations of systems, and diseases and diagnostics; and emergent macro-level biological and social phenomena. Analysis of cross-cluster dynamics shows how these domains integrated from the 1990s onward via technological and informatic capabilities — captured by MeSH belonging to the "Technology, Industry, and Agriculture" (J) and "Information Science" (L) branches — representing highly

* Corresponding author.

controllable, scalable and permutable research processes and invaluable imaging techniques for illuminating fundamental yet transformative structure–function–behavior questions. Our results indicate that 8.2% of biomedical research from 2000 to 2018 include MeSH terms from both the J and L MeSH branches, representing a 291% increase from 1980s levels. Article-level MeSH analysis further identifies the increasing prominence of cross-domain integration, and confirms a positive relationship between team size and topical diversity. Journal-level analysis reveals variable trends in topical diversity, suggesting that demand and appreciation for convergence science vary by scholarly community. Altogether, we develop a knowledge network framework that identifies the critical role of techno-informatic inputs as convergence bridges — or catalyzers of integration across distinct knowledge domains — as highlighted by the 1990s genomics revolution, and onward in contemporary brain, behavior and health science initiatives.

Keywords: Recombinant innovation; knowledge networks; convergence science; interdisciplinary research; team science.

1. Introduction

The codification of knowledge facilitates more efficient search across the vast space of possible creative inputs accessible to scientists [1] — conceived in research as strategic yet complex configurations of established and new entities, relationships, tools, equipment, methods, processes, observation, theory [2]. Organization of knowledge into a structured ontological map that specifies relationships between entities thereby facilitates the understanding of the structure, dynamics and future trajectory of knowledge [3–5]. Knowledge maps [6–9] are critical to managing the cognitive information load, as they improve the ability to establish and recall complex relationships [10]. Ontologies are also critical to the practical process of interdisciplinary integration, which requires scholars from different domains to become translators of both core and peripheral theory, thereby "effecting ontological transformation in the objects and relations of research" [11]. Widespread efforts to catalog entities beyond objects, i.e. to include higher-order concepts and relationships, is exemplified by the pervasive drive to assemble, organize, and cross-integrate high-resolution **-omics* atlases [12, 13]. These atlases are proliferating as tools that help scholars manage the increasing volume and rate of knowledge production [14, 15], and to accelerate breakthrough discovery [16–18] by helping scholars manage the uncertainty associated with exploration [19, 20].

Scholars of science have documented historical cycles of convergence and divergence [21, 22] responsible for conceptual churning that fuels convergent knowledge integration across distant originally distinct disciplines — as highlighted in the case of the genomics revolution [23] and human brain science [24]. Yet we lack a systematic knowledge-oriented perspective about the historical evolution of knowledge integration, as prior studies often apply inexact proxies for the knowledge and disciplinary content of research. For example, some approaches use journal-level research area categories as proxies for article-level discipline categories — an approach which mis-represents knowledge diversity at the article level, since this approach implies that all articles from the same journal are associated with the same (often a

single) disciplinary category.[a] Another common approach is using cited references as a proxy for the knowledge categories. Yet, again this approach has strong limitations because it uses the second-order knowledge (cited references) as a proxy for the knowledge contained in the original research, and still does not overcome the issue of journal-level categories being under-representative, and is also subject to biases deriving from idiosyncratic citing behavior of different authors. For further discussion of the differences between journal and article-level classifications see [18, 25]. In summary, the most appropriate approach to measure knowledge recombination and convergence is to use article-level keywords, which are independent of the publishing journal's classification, and more directly capture the first-order knowledge incorporated in the research itself.

To address the methodological limitations associated with journal-level disciplinary category systems, we developed a generalizable framework based upon the quasi-hierarchical Medical Subject Headings (MeSH) system developed and maintained by the US National Library of Medicine (NLM) [26, 27]. This ontology is comprised of roughly 30,000 standardized article-level keywords and is applied to the entire PubMed index maintained by the NLM. Hence, we combine both these data sources (PubMed and MeSH ontology) in comprehensive analysis of MeSH extracted from 21.6 million research articles published between 1970 and 2018 with the overall objective to systematically analyze the emergence of recombinant triple helix innovation as a source of biomedical convergence [28].

Importantly, the tree-like network structure of the MeSH ontology provides two complementary measures of cognitive distance between any pair of MeSH terms. The first measure is deduced from the 16 explicit MeSH branch categories that define different knowledge domains, such that if two MeSH belong to different knowledge network branches then they are categorically distinct. The second distance measure is defined according to the membership of MeSH within knowledge clusters, which we ascertain by applying network clustering methods to empirical MeSH–MeSH

[a]By way of example, all articles published by prominent multi-disciplinary journals *Nature*, *Science*, *PNAS*, *PLoS One* and Scientific Reports that are indexed by Web of Science are classified as "Multidisciplinary Sciences" according to the WC (Web of Science Category) field; and are classified as "Science Technology Other Topics" according to the SU (Research Area) field (also denoted as the SC field); and in the Scopus index they are classified as "MULT" according to the SUBJAREA field. Hence, WC, SU and SUBJAREA under-represent the knowledge diversity represented by individual articles published within this and other high-impact journals and megajournals of the same classification. As such, approaches to measure knowledge diversity based upon WC, SU and SUBJAREA systematically under-represent the diverse knowledge integrated within transformative research commonly published in these types of journals. A more nuanced issue is that measuring knowledge diversity based upon WC, SU or SUBJAREA of articles is confounded (in essence pre-determined) by the parent journal sizes, which pre-determine the variation of journal-level classification varieties, and consequently, provide a limited proxy for article-level knowledge diversity. This issue is not negligible in citation network analysis, as high-impact journals are over-represented as cited sources, and megajournals are over-represented as citing sources.

co-occurrence matrices. Afforded these operationalizations of a knowledge–knowledge metric distance, we address the following 3 research questions (RQs):

RQ1: To what degree have the rates of convergence — i.e. boundary-spanning knowledge recombination — shifted in biomedical research over the last half century?
RQ2: How and to what degree is biomedical convergence mediated by distinct over-represented MeSH — i.e. convergence bridges?
RQ3: What role have technological and informatic capabilities played in biomedical convergence, and how does this relate to other paradigms such as team science and convergence science, and how does it vary across different scholarly communities?

To address RQ1, we develop a statistical counting and visualization framework for identifying time-dependent knowledge clusters comprised of individual MeSH. These methods utilize the distinct branch structure of the MeSH ontology to define a measure of cross-domain knowledge integration, operationalized as variation and disparity of MeSH at the article level, thereby developing a high-resolution measure of recombinant innovation [1, 2] tailored for analyzing PubMed/MeSH. To address RQ2 we analyze the combination of distinct knowledge domains within individual research articles, which we quantify by measuring the interdisciplinary integration of knowledge across originally distinct domains according to the definition of *convergence science* proposed by the US National Research Council [18, 21, 23, 24, 29–33]. And finally, to address RQ3 we apply statistical analysis to sets of PubMed article grouped by year, team size and journal to identify robust trends and systematic variation.

Our results develop a comprehensive perspective on the emergence and impact of two related yet distinct domains of general purpose technology in biomedical research [34]. While our results highlight the well-appreciated values and implications of general purpose technologies [34], we further contribute to the literature by showing in historical and ontological detail how the emergence of techno-informatic capabilities as highly scalable and reusable research catalysts — such as non-invasive imaging, high-throughput anomaly detection, and big data integration — have transformed the scope of biology to meet the grand challenges at the frontiers of brain, behavior and health science [24].

2. Background and Literature Review

2.1. *Convergence — an intrepid interdisciplinary paradigm*

With the advent of synthetic biology and gene editing prowess in the 21st century, the scope and mission of the biomedical enterprise have transitioned from a descriptive origins towards more mechanistic understanding of the brain, behavior and health sciences, following a common pattern of disciplinary transformation away

from reductionalism and towards understanding causality and complexity [35]. Accordingly, in addition to defining biological elements and concepts and their relations, research recombines different forms of knowledge that are key to experimental approaches: tacit inputs on the one hand, the likes of which involve trial, error and practice (e.g. manual laboratory techniques); combined with explicit inputs on the other hand, which are highly codifiable, transmissible and therefore incrementally modifiable and scalable (e.g. pre-assembled computer algorithms that can readily be copied and altered).

Consequently, the theory of recombinant innovation [1, 2, 36] and innovation networks [37, 38] provide a powerful framework for understanding the potency, and uncertainty, embodied by combinatorial approaches to search, refinement, experiment and discovery. Likewise, the triple helix model of innovation [28, 39, 40] establishes the importance of catalysts that bring potent opportunities to fruition — i.e. challenging problems demanding novel solutions met with a diverse supply of research approaches [24].

The catalysts highlighted in what follows are techno-informatic capabilities, characterized by highly controllable, scalable and permutable research processes for exploring and testing the exponentially vast number of possible biological interactions and pathways. For this reason, we posit that biology was primed for convergence via the incorporation of high-throughput approaches to recombinant innovation.

To prime the pump for disciplinary convergence and to help address the excess risk and lower funding success associated with interdisciplinary research [41], national innovation systems have increased investment around transdisciplinary funding [42] by promoting a paradigm defined by its originators as "the coming together of insights and approaches from *originally distinct* fields" [18, 22, 31, 33] — as opposed to subfield integration characteristic of more narrow interdisciplinary endeavors [43–45]. The objective of convergence is thus to stimulate triple helix configurations that foster the emergence of both new hybrid disciplines [30] and specialized sub-disciplines [35] for addressing complex global challenges [18, 46]. Convergence is thus defined within the spectrum of interdisciplinarity [11] as far-reaching integration of distant knowledge domains into "new science" domains focused upon theories of causality and emergent complexity [35].

Recombinant innovation is also fundamental to the *convergence science* value proposition, as integrating diverse teams of experts hedges against uncertainty underlying the exploration process [47] and provides a testable mechanism [23, 24] for explaining the propensity of larger teams producing higher impact science [48]. Indeed, the potency of convergence was clearly evident in the genomics revolution, wherein traditional biologists and computer scientists leveraged familiar operational language (bits on the one hand, and base pairs on the other) to overcome relatively large epistemic and cultural distances between their traditionally distinct fields [23].

2.2. *Knowledge mapping and interdisciplinarity research*

Research on scientific ontologies and interdisciplinarity uses journal and article classification systems of varying granularity and contextuality. A common objective is to measure disciplinary diversity by using basic classification systems as proxies for research topic subject areas (SAs), and measuring category–category correlations [49, 50]. Some studies utilize broadly defined categories, such as the journal-level "Subject Category" descriptors implemented within Web of Science [51–53], or article-level systems used within multi-disciplinary journals such *PNAS* [54] and *PLoS One* [55]. However, the use of journal-specific classifications as the basis for interdisciplinarity measures comes with notable limitations, as we previously expounded.

Yet another stream of research utilizes more high-resolution article-level classification systems, such as keywords [45], International Patent Classification (IPC) and US Patent Office Classification (USPC) codes [1, 56–58], MeSH [9, 27, 28, 54, 59, 60], and Physics and Astronomy Classification Scheme (PACS) codes [44]. Other approaches include topic mapping based on word-frequency and co-word analysis [61, 62] and hybrid approaches integrating keywords and cited references to group research into knowledge clusters [8].

Given the non-uniform use of terminology and the varied methods for measuring interdisciplinarity [49], we start from its definition as the"bringing together distinctive components of two or more disciplines" [63]. Against this backdrop, convergence was conceived from the holistic perspective of mission-oriented national science policy [31] as the strategic (non-passive) integration of disciplines with distinct origins to achieve targeted objectives, e.g. research combining Humanities and Arts with STEM [64]. Such distances exacerbate preexisting cognitive and socio-buraucratic challenges associated with interdisciplinary research, in particular the problem of synergistically integrating and subsuming conceptual ontologies.

Another notable relation regarding convergence is its formulation as a facilitator of inventive forms not intended within the original logics of interdisciplinarity [11]. As such, convergence represents an intrepid form of interdisciplinarity in terms of the number, distance and novelty of the disciplinary configurations entailed [24, 63], that together foster holistic pathways towards understanding intractable phenomena, complex systems, and wicked problems [18, 32, 35, 65].

3. Data

3.1. *The MeSH tree*

We operationalize the representation of the biomedical knowledge network using article-level MeSH annotations, which can be considered as keywords with varying specificity that are interrelated within a 13-level quasi-hierarchical tree — i.e. an ontology. The official MeSH tree is maintained by the NLM and consists of 16

branches designated by the characters A, B, C, D, E, F, G, H, I, J, K, L, M, N, V, Z (see https://meshb.nlm.nih.gov/treeView to explore the tree). The six branches A, B, C, D, E, G shown in Fig. 1(a) represent core biological entities, concepts and methods. The branches F, J, L, N represent peripheral domains comprised also of entities, concepts and methods. To illustrate branch substructure, Fig. 1(b) shows the explicit MeSH–MeSH relations endowed in the MeSH tree for branch A ("Anatomy"). This particular branch is dominated by biophysical structures and their attributes — i.e. representing one half of the structure-function dichotomy, with branch G ("Phenomena and Processes") representing the other mechanism-oriented half.

While previous studies focused on MeSH belonging to specific branch subsets, such as C, D and E [27, 28], here we only exclude MeSH from the H, I, K, M, V, Z branches. We do not consider these MeSH branches because they are comprised of non-technical MeSH that label article characteristics or research contexts rather than relevant research inputs. For example, branch H classifies "Disciplines and Occupations" domains; V ("Publication Characteristics") classifies study characteristics and support of research, among other metadata; and branch M "Named Groups" classifies various subjects of human research, e.g. "Child, Orphaned" [M01.108]. We exclude branches K ("Humanities") and I ("Anthropology, Education, Sociology, and Social Phenomena") so that our measures of cross-domain convergence more appropriately identify configurations relevant to biomedical and health sciences rather than the social sciences.

3.2. *Article-level data*

Our comprehensive analysis is based upon the 2020 PubMed index consisting of more than 30 million index entries, with 93% of these classified as "Journal Article" or "Review". We then pruned this dataset of articles lacking Major MeSH and focused on the sample spanning 1970–2018, resulting in 21.6 million publications. Regarding notation, in what follows we use subscript p to indicate article-level information such as publication year, indicated by y_p; the number of coauthors, k_p; and the set of MeSH "keywords", represented as a vector $\mathbf{m}_p$ with 29,638 elements, each representing a distinct MeSH term. For example, an article with the sample average of 4 Major MeSH terms corresponds to $\sum_i \mathbf{m}_{p,i} = 4$.

3.3. *Projecting MeSH onto SAs*

We analyzed all MeSH belonging to the 10 core and peripheral branches. We refer to these broad knowledge domains as SAs, which represent a basis set for classifying MeSH according to their first-level parent branches (denoted by L_1). The MeSH tree is quasi-hierarchical (containing a relatively small number of loops), with 12 hierarchical levels per branch [27]. Hence, based upon this explicit relational tree structure, we can map any MeSH occurring at the third-level or greater to its

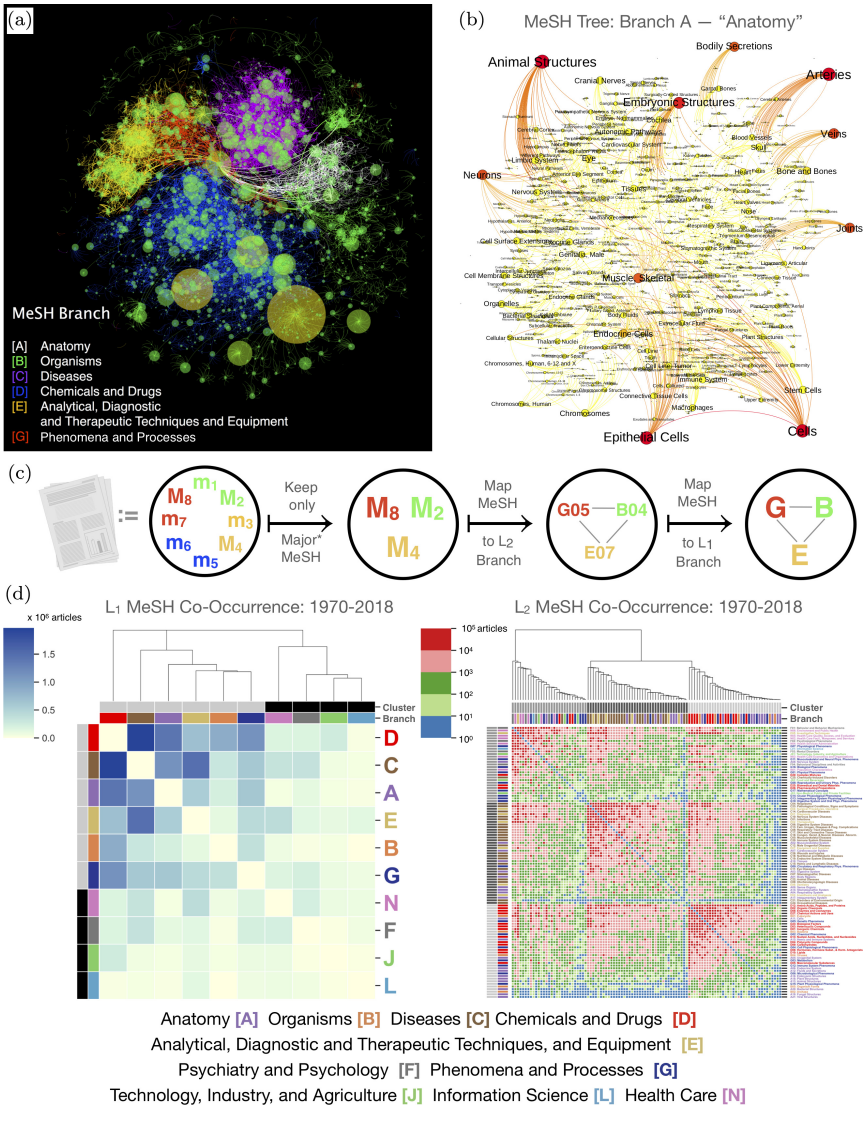

Fig. 1. Biomedical Knowledge Network. (a) Explicit network structure defined by the MeSH tree implemented by the US NLM within PubMed. Visualized is the quasi-hierarchical MeSH–MeSH network composed from the six traditional biomedical branches. (b) Network visualization of the "Anatomy" subtree (branch A). Nodes are individual MeSH and links are prescribed by the NLM MeSH Tree; nodes sized and shaded according to node degree. (c) Serial refinement of the MeSH appearing within each PubMed article: the set of MeSH $\mathbf{m}_p$ for each article p are refined to just the Major "keywords" (indicated within PubMed by an asterisk $*$); these Major MeSH $\mathbf{m}_p^*$ are readily mapped to their parent MeSH at the second level (L_2) and first level (L_1). (d) Historical MeSH co-occurrence frequencies for articles and reviews at the L_1 and L_2 levels; MeSH branches indicated by color-scale border segments (note the different color schema than in panel (a)). MeSH clusters are determined by a modularity maximizing algorithm [68] and indicated by the gray-scale border segments; hierarchical structure indicates the minimum spanning tree representation of each co-occurrence matrix; entities are sorted within clusters in decreasing order of prevalence, calculated as the sum of co-occurrences with all other MeSH.

corresponding second-level (L_2) MeSH term, for which there are 104 types.[b] See Appendix A.1 for further details.

Figure 1(c) illustrates the full process for producing $\mathbf{SA}_p$, which begins with first pruning out minor MeSH terms and auxiliary qualifiers, leaving just Major MeSH terms that represent the article's core SA decomposition. By way of example, the single MeSH term "Obesity" has four Tree Number locators corresponding to three SA (C, E and G): C18.654.726.500; C23.888.144.699.500; E01.370.600.115.100.160.120.699.500; G07.100.100.160.120.699.500. Hence, the L_2 representation of this single MeSH term is then given by {C18, C23, E01, G07} (note that these 4 L_2 codes are themselves distinct MeSH terms due to the tree structure). We catalog these MeSH terms using a L_2 count vector, denoted by $\mathbf{SA}^{(2)}$, which contains 104 elements. Further projecting this set to its L_1 representation yields {C, C, E, G}, represented by $\mathbf{SA}^{(1)} = \{0, 0, 2, 0, 1, 0, 1, 0, 0, 0\}$. This example highlights the nuanced organizational structure of the MeSH tree, wherein individual terms can represent multiple SA contexts; however, the frequency of these boundary-spanning MeSH are relatively low: 8% of all MeSH when considering all 16 branches; and just 6% of all MeSH when considering only the 10 focal SA analyzed here.

We then obtain a quantitative signature of each article's SA composition by combining all SA counts across all the MeSH associated with each article p into the aggregate vector $\mathbf{SA}_p$.

4. Results

4.1. *MeSH co-occurrence at the L_1 and L_2 levels*

We assessed the entire space of MeSH combinations by tallying SA co-occurrences among the nonzero elements contained in $\mathbf{SA}_p^{(1)}$ and $\mathbf{SA}_p^{(2)}$. Continuing with the example above, consider the SA counts {C, C, E, G}, which could arise from an article with a single Major MeSH annotation of "Obesity", or an article with four Major MeSH mapping individually to the same set of SA; we do not distinguish between these two cases, because the former example is highly unlikely. We then tabulate the SA co-occurrences by counting for all unique SA–SA dyad types. In such a case, we tabulate three co-occurrence types: CE, CG and EG. For related work analyzing the temporal evolution of all MeSH–MeSH combinations up to fourth order, see [2].

When aggregating co-occurrence tallies across articles, we combine normalized tallies — e.g. by assigning 1/3 weight to the CE, CG and EG matrix elements in the example above — so that each article contributes a total weight of 1 to the

[b] The longstanding MeSH ontology originates from the 1960s. Individual MeSH are assigned to articles by a professional team of U.S. NLM librarians based upon manual inspection of the full article text; the MeSH tree is also controlled and regularly updated by the NLM — for more details on how new MeSH terms are introduced, integrated into the MeSH tree, and back-filled, see [66, 67]. Notably, since we map all MeSH to their L_1 and L_2 representations, we do not expect our analysis to be incredibly sensitive to the introduction of new MeSH terms, which tend to be specific concepts added at the higher-level MeSH-network periphery.

co-occurrence matrix $\mathbf{M}^{(1)}$ (and similarly for $\mathbf{M}^{(2)}$ constructed at L_2). In this way, the total sum across matrix elements of $\mathbf{M}$ is proportional to the total number of articles analyzed in the sample.

To begin, Fig. 1(d) shows co-occurrence frequencies recorded in the symmetric matrices $\mathbf{M}^{(1)}_{1970-2018}$ and $\mathbf{M}^{(2)}_{1970-2018}$, calculated across articles published between 1970 and 2018. Each matrix indicates the SA name along the right border, accompanied by a consistent color indicating the corresponding L_1 branch, useful for guiding visual inspection. We grouped SA into knowledge clusters by applying the weighted version of the Louvain modularity maximizing algorithm [68], which accounts for the fact that most if not all of the $\mathbf{M}^{(1)}$ and $\mathbf{M}^{(2)}$ matrix elements are nonzero. The resulting knowledge clusters are indicated by the gray-scale border segments along the upper and left matrix borders. The top border contains hierarchical clustering trees indicating the minimum spanning tree representation of each matrix. To further aid visual inspection, we maintain the gray-scale cluster shades across the aggregate matrices shown in Fig. 1 as well as the time-disaggregated matrices visualized in Fig. 2. The time-aggregated visualization, which facilitates the identification of individual MeSH, is intended to serve as a primer for the time-disaggregated visualizations, which instead focuses on communicating the dynamics of the knowledge clusters.

As such, we analyzed MeSH co-occurrences at both the L_1 and L_2 levels, thereby producing knowledge network maps with complementary degrees of granularity. The main result for L_1 is the two-SA substructure observed for $\mathbf{M}^{(1)}_{1970-2018}$. This substructure confirms that branches A, B, C, D, E and G form a core biomedical cluster, with the peripheral application domains N, F, J and L forming a second cluster. Figure 1(a) shows the MeSH–MeSH network for MeSH belonging to this core biological SA.

At higher resolution, the L_2 matrix $\mathbf{M}^{(2)}_{1970-2018}$ features three mixed clusters. The first cluster is comprised of a wide array of L_2 MeSH pertaining to complex human phenomena related to behavior, physiology and health. Importantly, this cluster also includes L01 ("Information Science") and J01 ("Technology, Industry, and Agriculture"). Because these are two focal domains in our analysis, we refer to their MeSH scope notes to provide additional context. According to MeSH scope notes, "Technology" refers to "the science and application of techniques" (J01), and more specifically "the application of scientific knowledge to practical purposes in any field [including] methods, techniques, and instrumentation" (J01.897). Likewise, MeSH scope notes describe "Information Science" (L01) as "The field of knowledge, theory, and technology dealing with the collection of facts and figures, and the processes and methods involved in their manipulation, storage, dissemination, publication, and retrieval."

The second cluster is comprised of diagnostic methods associated with particular diseases and the systems they affect, as represented by MeSH primarily from branches A ("Anatomy"), C ("Chemicals and Drugs") and E ("Analytical, Diagnostic and Therapeutic Techniques, and Equipment"). And the third cluster is comprised

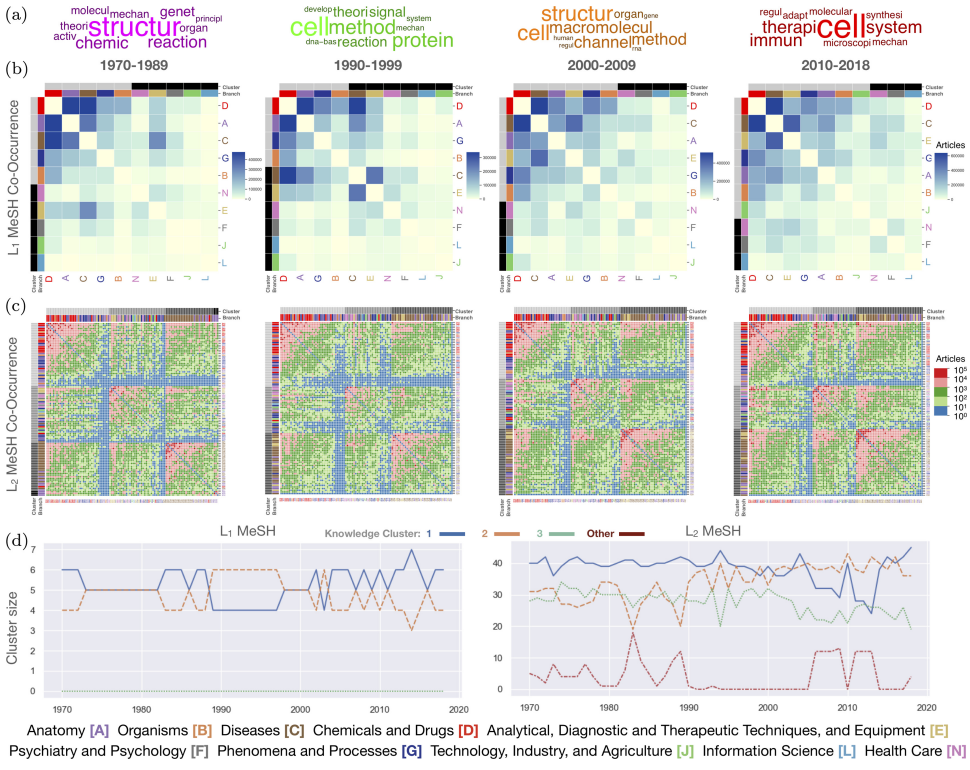

Fig. 2. Structure and Dynamics of MeSH–MeSH Co-occurrence. (a) Top 10 most frequent destemmed keywords from Chemistry and Physiology or Medicine Nobel Prize rational statements, sized proportional to their prevalence, illustrating the transition from structural to mechanistic orientation of grand scientific pursuits. (b, c) MeSH branches are indicated by axis labels with corresponding color-scale border segments; MeSH clusters determined by the Louvain modularity maximizing algorithm [68] are indicated by the outer gray-scale border segments. MeSH are ordered within clusters in decreasing order of prominence corresponding to total number of article instances for a given period, as indicated by the numerical scales. (b) L_1 co-occurrence matrices show sequential dynamics across 4 non-overlapping periods; linear color-scale. (c) L_2 co-occurrence; logarithmic color-scale. (d) Co-occurrence cluster size dynamics by year and resolution level: L_1 (left) and L_2 (right).

of biological entities and emergent phenomena, from cells to chemicals and the multi-scale processes that connect inputs, outputs and characteristics of their reaction environments.

4.2. *Historical co-occurrence trends*

The last half-century of biomedical research has witnessed incredible transition from a descriptive, reductionist field [35] — focused on identification of molecules, higher-order structures, reaction pathways and mediators — into a field seeking to identify holistic mechanisms underlying systemic abnormalities in an effort to develop pointed therapies. This transition is illustrated through the Nobel Prizes for

Chemistry and Physiology or Medicine, which were predominantly awarded for research documenting key structural entities and reactions up until the 1970s and 1980s. Building upon this early work, subsequent Nobel research has transitioned to addressing challenges at the nexus of the medical innovation triple helix — where societal demand and industrial supply of acute solutions are increasingly mediated by techno-informatic capabilities [28, 69].

To illustrate this point, Fig. 2(a) shows the top ten stemmed keywords from Nobel Prize "rationale" statements, identifying prominent contextual themes across the last half century. Indeed the 1980s brought forth novel imaging, laboratory control and synthesis technologies, fundamental in the 1990s to map the genetic blueprints encoding biological structure [23]. Polymerase chain reaction (PCR), among other high-throughput techniques, accelerated science to the realm of highly controllable, scalable and permutable processes that hitherto have been primarily restricted to traditional computational domains. Since then, the last twenty years has been marked by the rapid development of bioengineering capabilities, such as CRISPR gene editing tools [70] and powerful demonstration of stem cell pluripotency [71], which allow scientists to more precisely understand emergent structure-function relations, opening the door for synthetic biology applications [72] that (re)design biological systems [73] or even develop altogether new building blocks [74].

To visualize this history unfold, we disaggregated the MeSH co-occurrence data into four non-overlapping periods: 1970–1989, 1990–1999, 2000–2009 and 2010–2018. For each period we tabulate the co-occurrence matrix, denoted by $\mathbf{M}_y^{(1)}$ (or $\mathbf{M}_y^{(2)}$), where y denotes a given period. Figure 2(b) illustrates the structural dynamics at the L_1 level, identifying branch E and J as vacillating domains switching between the two dominant clusters, which are further explored in the next section on knowledge network reorganization. The first L_1 cluster is characterized by D, A, G and B — largely descriptive SA, with the exception of "Phenomena and Processes" (G). The second cluster is comprised of N, F and L — peripheral SA which are increasingly convergent with the traditional SA, as indicated by extreme off-diagonal elements representing cross-domain integration.

At higher resolution, Fig. 2(c) shows the structural evolution of $\mathbf{M}_y^{(2)}$, which is characterized by two relatively stable clusters and a more diverse and vacillating intermediate cluster. Note that MeSH are sorted within each cluster in decreasing order of prevalence, calculated as the sum of co-occurrence values across a given row of $\mathbf{M}_y^{(2)}$. Notably, "Information Science" (L01) and "Technology, Industry, and Agriculture" (J01) are located in this intermediate cluster during the 1970–1989 period, where they play less dominant roles as indicated by their ranks within the cluster. However, over time their prominence within this cluster increases; by 2010–2018, J01 joined the biological agents and reactions cluster, characterized by the core "Amino Acids, Peptides, and Proteins" (D12), "Chemical Actions and Uses" (D27), "Eukaryota" (B01) and "Organic Chemicals" (D02). Close inspection reveals J01 having relatively strong co-occurrence with "Inorganic Chemicals" (D01),

"Investigative Techniques" (E05), "Health Care Quality, Access, and Evaluation" (N06) and "Biomedical and Dental Materials" (D25) — indicative of a highly convergent applications nexus facilitated by technological capabilities.

Close inspection further reveals a similar transformation for L01, which rose to prominence from the 1980s onward, as genomics and other panoramic *omics* revolutions increased the demand for informatic solutions to map, characterize and associate entities into a functional atlas. This role did not merely manifest from increased technological capabilities associated with "Diagnosis" (E01) and "Investigative Techniques" (E05), but also involved the integration of powerful "Mathematical Concepts" (G17), in particular "Algorithms" (G17.035) to optimize classification and AI methods for "Deep Learning" (G17.485.500). Such methods are critical for capitalizing on new data-driven opportunities in Health Care (represented broadly by N04, N05, N06) for understanding "Behavior and Behavior Mechanisms" (F01).

5. MeSH Cluster Dynamics Indicate Periods of Knowledge Network Reorganization

The knowledge network dynamics, in particular bursts of cluster size variation exhibited in Fig. 2(d), allude to paradigm shifts mediated by innovation [2] and innovator dynamics [65, 75, 76]. To identify turbulent periods in the knowledge network, we developed a network method for identifying fluctuations of individual MeSH constituents across clusters by tracking the entry and exit of constituents between sequential 1-year periods.

The motivation for the following clique-labeling method is the following — how do we differentiate between individual MeSH (denoted by m) changing clusters when the clusters themselves are dynamic? The solution that we introduce is a time-invariant labeling system for uniquely identifying a given clusters C_x, independent of its exact composition. To achieve this, we rely on a fixed constant within the co-occurrence network dynamics, namely those MeSH groups that are always found clustered together, independent of t. We label these MeSH groups as cliques, denoted by q_n. This dimensional reduction approach shifts the aforementioned problem to instead evaluating to what degree the set of cliques (denoted by $Q_{m,t}$) within the cluster that houses a given m in a given period t are (dis)similar to those in the previous period $t-1$.

Figure 3(a) shows the basis set of MeSH cliques, q_n, where $n = 2$ for L_1 and $n = 14$ for L_2. An interesting cross-SA clique observed at the L_2 level is the dyad formed by "Technology, Industry, and Agriculture" (J01) and "Biomedical and Dental Materials" (D25). By contradistinction, we also observe a mono-SA clique formed by seven L_2 branches (D01, D02, D03, D04, D09, D10, D27), all members of the L_1 D-branch. Hence, these stationary cliques are motifs that uniquely identify an arbitrary cluster C_x according to its constituent cliques.

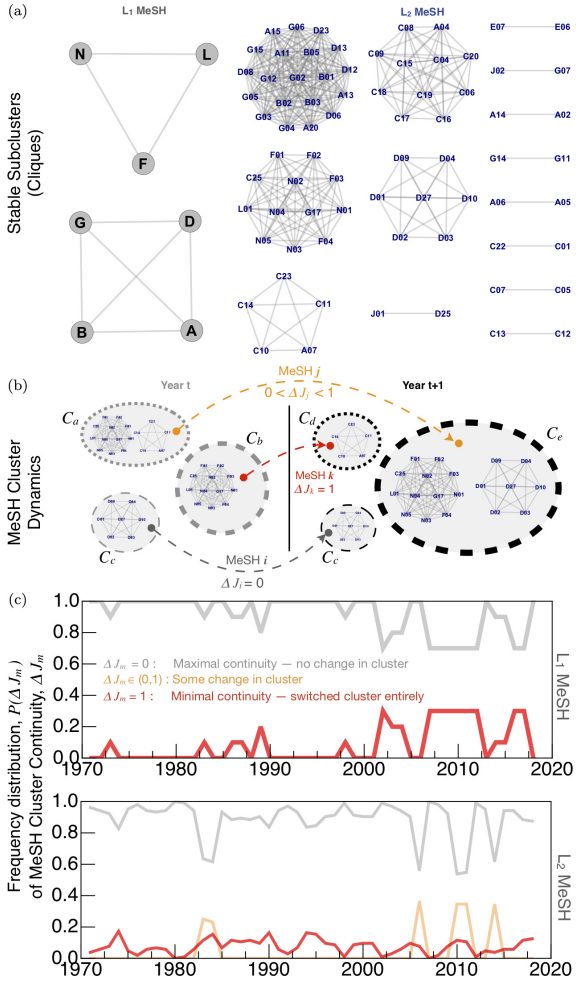

Fig. 3. (Color online) Method for quantifying fluctuations of individual MeSH across knowledge clusters. Reorganization of co-occurrence networks are indicative of paradigmatic disruption and cross-domain bridge formation. Stable subclusters are groups of MeSH that always appear together in annual-level co-occurrence matrix Louvain clustering — see Fig. 2(c). These cliques are used as a unique identification system for calculating the cluster continuity ΔJ_m of a given MeSH term m over sequential 1-year periods. (a) Cliques for the period 1970–2018 calculated for the L_1 (left) and L_2 (right) levels. (b) Schematic of the identification system for tracking cluster dynamics of individual m. For example, comparing MeSH i for the sequential year t and $t+1$ periods, i stayed in the cluster C_c (identified as having the D-clique comprised of seven L_2 MeSH from the D-branch) and so there is maximal continuity of the cliques defining its surrounding cluster. Contrariwise, MeSH k transitioned from a cluster in t defined by a single clique that differs from the clique defining its cluster in $t+1$; hence, this case corresponds to minimal continuity. The case of MeSH j is inbetween these extreme cases, whereby j transitions from a cluster in t that shares a common clique as the cluster in $t+1$, but with differing second member clique. (c) Frequency distribution of continuity values ΔJ_m by year. Years with low MeSH cluster continuity feature higher frequencies of MeSH that switch knowledge clusters entirely (red) or partially (orange); conversely, periods with high cluster continuity, in which all m stay in the same cluster, corresponds to the gray curve approaching unity. The period with the highest levels of knowledge cluster discontinuity started in the early-2000s and peaked around 2010. Discontinuity peaks indicate the emergence of individual MeSH serving as cross-cluster bridges.

By way of example, Fig. 3(b) shows the cluster C_c which contains the D-clique described above in two subsequent years. Consequently, an arbitrary MeSH i belonging to C_c in year t and $t+1$ is stationary with respect to the cliques. Alternatively, consider an arbitrary MeSH j which was a member of C_a (identified by two particular cliques) and C_e in the next year (identified by two cliques, one from C_a and one altogether new one). In this case, the MeSH j (indicated by orange) has partially switched clusters. Finally, consider an arbitrary MeSH k (indicated by red) which was a member of C_b (identified by a single clique) and C_d in the next year (identified by a completely different clique); this case corresponds to minimal continuity (maximal discontinuity) with respect to cluster member cliques.

To quantify aggregate cluster dynamics, we assign each MeSH (represented generically by the index m) a set of labels $Q_{m,t} = \{q_1, \ldots, q_n\}$ corresponding to each of the n unique cliques present, as indicated by q, for a given year t. We then calculate the Jaccard distance $\Delta J_m = 1 - |Q_{m,t} \cap Q_{m,t+1}|/|Q_{m,t} \cup Q_{m,t+1}|$. Hence, $\Delta J_m = 0$ corresponds to maximal continuity (since $|Q_{m,t} \cap Q_{m,t+1}| = |Q_{m,t} \cup Q_{m,t+1}|$). Conversely, $\Delta J_m = 1$ corresponds to maximal discontinuity (since $|Q_{m,t} \cap Q_{m,t+1}| = 0$). When there is a partial shift in cluster member cliques, then we obtain intermediate values, $0 < \Delta J_m < 1$.

Figure 3(c) indicates the periods with the highest levels of MeSH network reorganization, characterized by $\Delta J_m > 0$ values (indicated by orange and red curves). These results are consistent for both the L_1 and L_2 resolutions. MeSH data aggregated at the L_1 level indicates a period of heightened MeSH cluster dynamics starting in roughly 2000 and continuing in bursts up to present. Additional inspection at the L_2 level indicates periodic fluctuations occurring over time, but again with heightened turbulence in the early 1980s, and in the years around 2010. We interpret these heightened periods of MeSH cluster discontinuity as indicators of individual MeSH that emerge as cross-cluster knowledge bridges that integrate distinct knowledge domains, a hallmark of biomedical convergence in topical SA space. In summary, the objective of this method is to identify particularly turbulent periods in the knowledge network. This provides a consistency check for results in the following sections, where we identify particular knowledge domains (represented by individual MeSH) that were central to the emergence of biomedical convergence.

5.1. *Biomedical convergence identified by the emergence of cross-cluster bridges*

To identify particularly important MeSH that served as bridges linking different knowledge clusters, we developed a metric that quantifies the degree to which individual nodes connect distinct clusters in a weighted and non-sparse co-occurrence network. This metric is motivated by the more sophisticated bridge centrality index [77], as well as the analysis in the previous section regarding cross-cluster MeSH dynamics indicating periods of knowledge network reorganization. We develop this

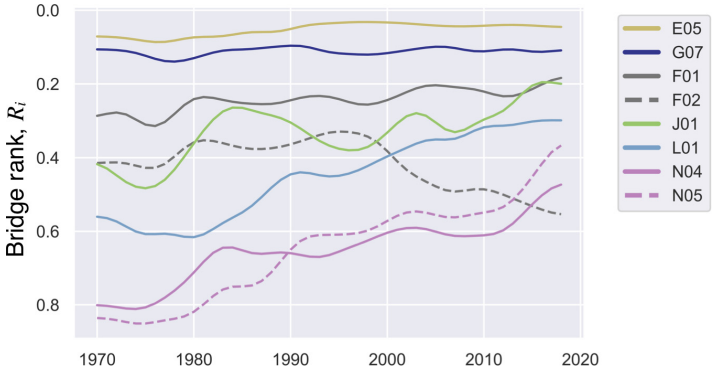

Fig. 4. Prominent convergence bridges. (A) For each MeSH, we analyzed the time series of bridge rank, $R_{i,t}$, a normalized ranking based upon the knowledge bridge score (β_i) defined in Eq. (A.1). Smaller rank values ($R_{i,t}$) correspond to larger β_i values, representing MeSH that are highly co-occurrent with MeSH belonging to other knowledge clusters. Plotted are smoothed time series to facilitate visual inspection. Notably, J01 and L01 are rapidly emerging convergence bridges arising from the highly generalizable, scalable and codifiable nature of techno-informatic tools and algorithms facilitating novel non-invasive imaging, high-throughput analysis, measurement, and data integration. Other MeSH shown here identify the emerging convergence nexus of Health Care (N04, N05) and Behavior (F01).

approach with $\mathbf{M}_t^{(2)}$ in mind, in which there are no nonzero matrix elements. Applying this method to $\mathbf{M}_t^{(2)}$ yields a knowledge bridge score for each MeSH and year, denoted by $\beta_{i,t}$, which is large if a particular MeSH is highly connected to clusters other than its own.

Figure 4 shows the rank dynamics $R_{i,t}$ (where larger $\beta_{i,t}$ correspond to smaller rank value R_i) for 8 MeSH groups that are distinguished by either persistent growth or sustained prominence as knowledge bridges over the entire study period. Notably, L01 features steady growth following its emergence in the early 1980s, whereas J01 features an oscillating upward trend. We also observe a divergence between "Behavior and Behavior Mechanisms" (F01) and "Psychological Phenomena" (F02), with the latter becoming increasingly insular over the last 20 years. The emergence of health care bridges relating to "Health Services Administration" (N04), "Health Care Quality, Access, and Evaluation" (N05) and situational behavior analysis (F01) is another disruptive trend from the last decade — consistent with the 2013 ramp-up of several international *Human Brain Projects* [24, 78, 79], thereby championing the brain research nexus as a convergence frontier [24].

5.2. *Techno-informatic capabilities facilitate biomedical convergence around brain, behavior and health science*

Concurrent co-occurrence of MeSH pairs (denoted by j and j') that are close neighbors of a given MeSH (i.e. the ego node, i) is indicative of recombinant knowledge domains mediated by i. To analyze these triadic closure phenomena as they relate to cross-domain convergence, we explored dynamic patterns occurring in the subset of $\mathbf{M}_{jj'}^{(2)}$ values among the most prominent co-occurring neighbors of a

given MeSH. In particular, we focused on the co-occurrence dynamics for each of the convergence bridges identified in Fig. 4.

For these 8 MeSH groups, we selected the ten most frequently co-occurring MeSH for the denoted period. Figure 5 shows the subset of $\mathbf{M}^{(2)}$ values visualized as a subnetwork. For a given period, the top ten co-occurring neighbors (j) are sorted in clockwise fashion starting from the top, with nodes sized proportional to $\mathbf{M}^{(2)}_{ij}$. To facilitate visual inspection, the SA of each MeSH is indicated by its node/label color. Links are plotted with thickness and shade proportional to $\mathbf{M}^{(2)}_{jj'}$, thereby indicating the cross-domain linkages among prominent neighbors that are facilitated by i. Each node includes its MeSH identifier and a knowledge cluster identifier, the latter indicated by a gray-scale gradient.

Many of the neighborhood subnetworks feature J01 and L01, together reinforcing the observation that techno-informatic capabilities facilitated biomedical convergence around emergent frontier of brain, behavior and health science. The most significant feature of each convergence bridge are as follows: "Investigative Techniques" (E05) integrate all clusters with high SA diversity. "Behavior and Behavior Mechanisms" (F01) integrates health, "Physiological Phenomena", "Eukaryotes" and "Pathological Conditions, Signs and Symptoms" domains. And F01 and "Psychological Phenomena" (F02) increasingly incorporate "Information Science" (L01) and "Investigative Techniques" (E05).

Representing the increasingly lucrative domain of human health science, "Physiological Phenomena" (G07) exhibits a high co-occurrence with "Food and Beverages" research, which also exhibits high diversity of cross-SA and cross-cluster. Highlighting the role of academic-industry-government cross-sectoral triple helix [39, 40], "Technology, Industry, and Agriculture" (J01) increasingly integrated from the 1990s onward with the "Chemical and Drugs" domain, thereby capturing the pharmaceutical industry. L01 integrates with "Mathematical Concepts" to facilitate research investigating "Behavior and Behavior Mechanisms" by leveraging technology providing non-invasive "Investigative Techniques". And both "Health Services Administration" (N04) and "Health Care Quality, Access, and Evaluation" (N05), which have become highly coupled mirrors of each other, have in the last decade incorporated "Information Science" methods for clinical diagnosis and classification of "Pathological Conditions, Signs and Symptoms" and overall health care program assessment.

5.3. *Convergence factors — SA composition, prevalence, team size and scientific impact*

The value proposition envisioned by convergence science originators [31] was novel configurations of expertise strategically assembled to address a dimension of the underlying problem that would be otherwise inaccessible to mono-disciplinary approaches [18]. The convergent union between biology and computing experts in the genomics revolution provides a rich example [23]. Hence, it follows that

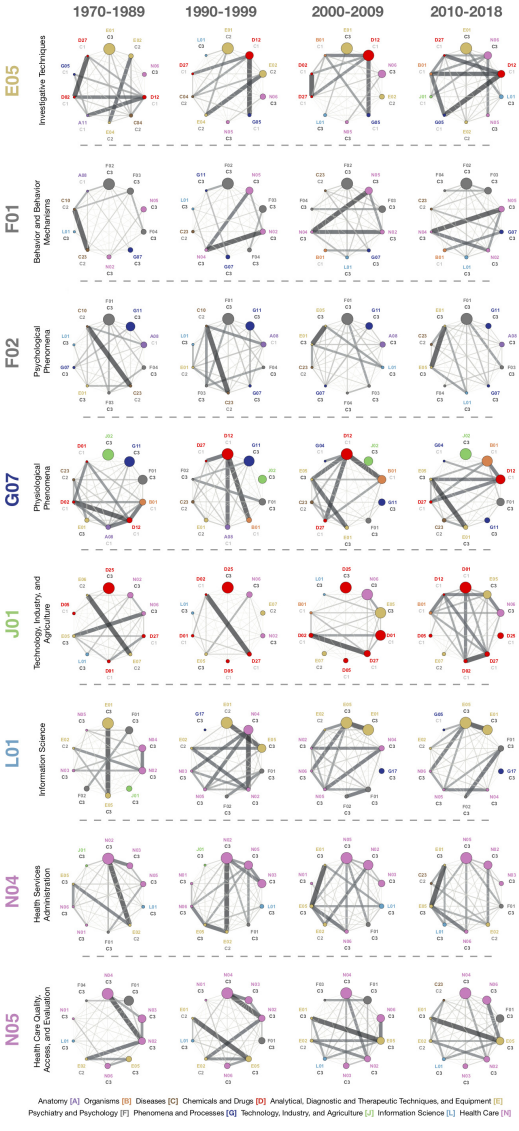

Fig. 5. Biomedical convergence around brain, behavior and health science mediated by techno-informatics bridges. Each row illustrates a MeSH term (i) identified as a convergence bridge. Each network is calculated using data from the indicated period, and shows the ten most frequently co-occurring MeSH — sorted clockwise, starting from the top, with nodes sized proportional to $\mathbf{M}_{ij}^{(2)}$; each MeSH's SA is indicated by its node/label color. Links are plotted with thickness and shade proportional to $\mathbf{M}_{jj'}^{(2)}$, thereby indicating the cross-domain linkages among prominent neighbors that are facilitated by i. Each node includes its MeSH identifier and a knowledge cluster identifier, the latter indicated by a gray-scale gradient. For example, MeSH J01 "Technology, Industry, and Agriculture" (which is a member of C3 for the first three periods and subsequently transitioning to C1 in the most recent period) is highly connected to MeSH from all other clusters (C1–C3), in particular to L01 until its disassociation in the most recent period 2010–2018; interestingly, L01 "Information Science" diverged from J01 as early as the first period 1970–1989, subsequently becoming more strongly coupled with members of branch E and G.

convergent research strategies are more likely to prevail in cross-disciplinary teams, thereby providing a testable mechanism confirmed in [23, 24] that explains the the propensity for larger teams to deliver high-impact science [48].

Indeed, the importance of informatic (L) and technological (J) capabilities to modern biomedical research cannot be understated. To illustrate the emergence of these X-disciplinary bio-technological and bio-informatic modes, indicated in short by X, we calculated the fraction of articles by year that contain a significant component belonging to SAs J and/or L. To ensure that the articles are otherwise focused on traditional biomedical SA, we estimate the number of cross-domain articles by focusing on the subset of PubMed articles featuring a majority (i.e. half or more) of their MeSH in the core categories ABCDEG. We then assign the indicator X to those articles that also contain at least a quarter of their MeSH belonging to L, J or L + J in combination.

Figure 6(a) shows the frequency of articles featuring X, calculated as the fraction $f_X(t)$ of the total articles by year, indicating a burst of activity around the early 1990s for research containing both L and J in combination — much greater than the $f_X(t)$ calculated for either J or L alone, thereby indicative of their complementarity as opposed to substitutability. In other words, they are observed far more frequently in tandem than individually. This particular configuration of ABCDEG × J × L represents a formidable nexus featuring the combination of high-throughput equipment to produce and churn through massive biomedical data. For comparison, Fig. 6(b) applies the same method to all MeSH (i.e. not distinguishing between Major and Minor MeSH), thereby including the more peripheral MeSH capturing the idiosyncratic research details. This second perspective indicates a more steady integration over time of J and L capabilities at the SA periphery, with the strongest upturn in J + L in the early 2000s, which has since saturated. Hence, the 1990s brought forth the revolution in genomics research, and the 2000s witnessed further penetration of this paradigm shift into other biomedical arenas, albeit their appears to be a recent stagnation. While it is beyond the scope of our analysis to identify the cause of this stagnation, we posit that it represents the saturation point, above which research is dominantly computational, and such research articles are not necessarily included in PubMed, which is biomedical oriented.

To further explore the evolution of cross-domain SA configurations at the article level, we developed a Blau-like diversity measure based upon SA–SA co-occurrence, denoted by $f_{D,p}$. This measure accounts for two distinct diversity types — both categorical variation and concentration disparity [80]. Our method takes as input the categorical count vector $\mathbf{SA}_p$ and applies the outer tensor product, $\mathbf{SA}_p \otimes \mathbf{SA}_p$, yielding a weighted matrix $\mathbf{D}_p$ that captures dyadic SA–SA co-occurrence; see Appendix A.3, in particular, Eqs. (A.2) and (A.3) for further elaboration. To summarize, if an article's MeSH descriptors are contained in just a single SA, then there will be only one single nonzero value contained in $\mathbf{D}_p$, which will occur along the matrix diagonal. Conversely, if the article features several SA, then off-diagonal elements quantify combinations of cross-domain SA as a weighted product. Hence,

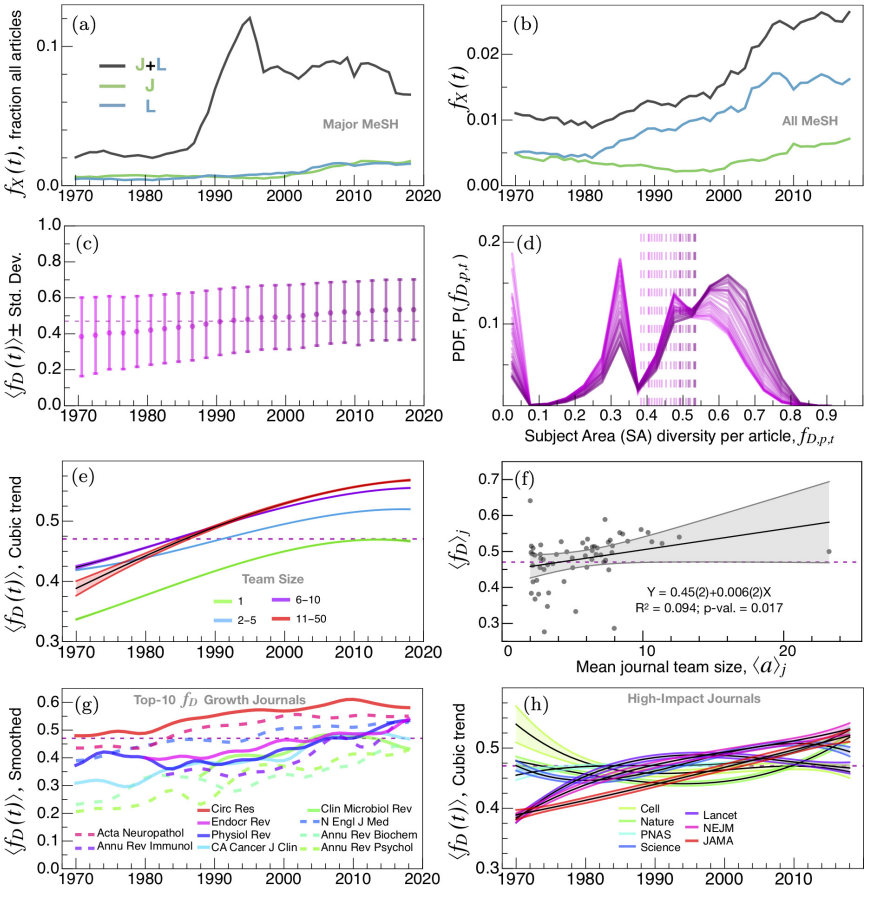

Fig. 6. Increasing prevalence of convergence science. (a) Fraction of articles, $f_X(t)$, featuring at least half of Major MeSH in ABCDEG and at least a quarter in J, L or J + L. (b) $f_X(t)$ calculated using all MeSH. (c) Average and standard deviation (error bar) of $f_{D,p}$, calculated for non-overlapping 3-year windows over the period 1970–2018; the global average diversity value (0.47) is indicated by the horizontal dashed line. The mean value $\langle f_D(t)\rangle$ has increased from 0.38 to 0.53, representing 39.5% growth, over the entire 49-year sample period. (d) Probability distribution $P(f_{D,p})$ for each of the 3-year subsamples analyzed in panel (C). One notable shift is the decreased prevalence of mono-domain articles ($f_{D,p} = 0$) which is a large but not sole contributor to the increasing trend in average value, $\langle f_D(t)\rangle$ (shown as horizontal vertical dashed lines). The right tail of the distribution is well defined and does not increase dramatically in range (i.e. the max diversity value is persistently around the value 0.9). The peak around value 1/3 (1/2) corresponds to articles with MeSH contributing equally to two (three) distinct SA and represents the first (second) mode of cross-domain convergence. (e) Increased knowledge diversity correlates with the emergence of team science, where the most rapid increase is among the larger teams with 11–50 coauthors. Solo-author research has only recently reached the average diversity value indicated by the horizontal dashed line. (f) Positive relationship between average team size and cross-domain SA diversity calculated at the journal level (ANOVA p-val. = 0.017; shaded region is 99% CI). (g) Ten biomedical journals featuring the highest growth in average $f_{D,p}$ values over 1970–2018. (h) Heterogenous trends by longstanding core (*Cell*), elite multi-disciplinary (*Nature, PNAS, Science*) and biomedical (*Lancet, NEJM, JAMA*) journals. *NEJM* and *JAMA* show the largest consistent increase over time; *Nature* and *PNAS* exhibit the largest increase in the last decade. Each curve represents the third-order polynomial trend, with shaded colored areas indicating the 99% CI for each trend line.

$f_{D,p}$ measures the relative fraction of off-diagonal to diagonal elements — with $f_{D,p} = 0$ corresponding to minimal diversity. Because more evenly distributed SA counts will yield relatively larger off-diagonal values, and hence larger $f_{D,p}$ values, it is both a measure of variation and disparity [80]. Moreover, $f_{D,p} \in [0, 1)$ is a bounded statistic with clear interpretation of the lower and upper bounds.

We computed $f_{D,p}$ based upon $\mathbf{SA}_p^{(2)}$ counts tallied for each publications in our PubMed sample (i.e. using the L_2 MeSH representation of an article). Figure 6(c) shows the evolution of the mean diversity value $\langle f_D(t) \rangle$ calculated for articles grouped by publication year, and shows a steady increase in SA diversity over the last half century. Analysis of the full distribution of values, denoted by $P(f_{D,p,t})$, also exhibits a systematic positive shift across the range of $f_{D,p}$ values, and so the increase in the mean value is not just the result of an upper tail effect in the aggregate distribution. Contrariwise, we observe a marked decrease in the prevalence of mono-SA articles characterized by $f_{D,p} = 0$ in the lower distribution tail — see Fig. 6(d). This result — the disappearance of mono-disciplinary research — is analog to the reduced frequency of single-authored research observed over the same period [48], indicative of the intense burden to integrate distant knowledge domains, a challenge that contributes to the disappearance of the "renaissance" solo genius [81, 82].

To assess how team size mediates article-level SA diversity, we disaggregated the data into four team-size groups: solo-author (coauthor number $a_p = 1$); small-sized team ($1 < a_p \leq 5$); medium team ($5 < a_p \leq 10$); and large team ($10 < a_p \leq 50$). The thresholds distinguishing the ranges for each team-size group were arbitrarily chosen, and roughly correspond to exponentially increasing bin sizes. One limitation of this approach, however, is that it does not account for the differing characteristic team sizes and their frequencies across discipline, nor how team sizes differentially mediate outcomes in different disciplines [83, 84].

Figure 6(e) shows the dominant trends in average diversity time series $\langle f_D(t) \rangle$ for each team size group, from 1970 to present. To facilitate visual comparison of the dominant trends, shown are each time series fit using a third-order polynomial: $\langle f_D(t) \rangle \equiv a + b(t - 1990) + c(t - 1990)^2 + d(t - 1990)^3$. We choose to terminate the nonlinear fit at third order so to be able to capture the growth saturation indicated at the aggregate in Fig. 6(c); And we center the time variable around 1990 in order to balance the fit around the temporal midpoint of our data sample to improve model fit quality.

Notably, the curve calculated for solo-author teams shows a significant systematic offset towards lower SA diversity values. From the mid-1980s onward, the curves are ordered according to team size group, with the largest teams achieving SA diversity levels far in excess of the unconditional average for the entire period, $\langle f_D \rangle = 0.47$ (indicated by the horizontal dashed line in each figure sub-panel). As a robustness check, Fig. 6(f) shows the journal-level relation between article-level SA diversity and mean journal team size for the 60 biomedical journals appearing in the set of top-100 journals ranked by 2018 Clarivate Analytics JCR Impact Factor, indicating a higher range of SA diversity associated with larger teams.

Table 1. Top-20 convergence journals. Biomedical journals in the top-100 2018 JCR Impact Factor, ranked according to average categorical SA diversity over the period 1970–2018

Journal (j)	SA diversity	
	Average $\langle f_D \rangle_j$	Std. Dev. $\sigma[f_D]_j$
Circulation Research	0.56	0.16
Journal of Clinical Oncology	0.55	0.14
Journal of American College of Cardiology	0.51	0.17
Acta Neuropathologica	0.51	0.15
Nature Genetics	0.51	0.16
Gut	0.49	0.17
Proceedings of the National Academy of Sciences	0.48	0.18
Science	0.48	0.18
Annals of Internal Medicine	0.47	0.19
New England Journal of Medicine	0.47	0.19
Cell	0.47	0.17
Nature	0.46	0.18
European Heart Journal	0.46	0.18
BMJ	0.46	0.20
Endocrine Reviews	0.46	0.19
Lancet	0.45	0.20
Clinical Microbiology Reviews	0.44	0.18
Physiological Reviews	0.43	0.21
Annual Review of Immunology	0.40	0.21
CA: A Cancer Journal for Clinicians	0.39	0.21

The wide variation in journal level $\langle f_D \rangle_j$ indicates that support for convergence science is a feature of the distinct communities of expertise. While it is beyond the present study, we speculate that the multidisciplinary composition of a journal's editorial board is likely to be a strong factor underlying the prevalence of convergence science featured in a particular journal. Among the most prestigious journals shown in Fig. 6(h), several feature recent decline in $\langle f_D(t) \rangle$ over the last two decades, including *Cell* and *Lancet*; contrariwise, the medical journals *NEJM* and *JAMA* are consistently trending upwards. To further illustrate the dynamics at the journal level, Fig. 6(g) shows the top-10 journals in terms of their largest growth in $\langle f_D(t) \rangle$ over the study period, featuring prominent journals in multiple domain areas including cancer, microbiology, medicine and psychology. Despite such consistent growth trends for some journals, this is not a universal feature. See Table 1 for prominent journals ranked according to their average SA diversity value $\langle f_D \rangle_{j,1970-2018}$.

6. Discussion

With each new discovery — some large, others incremental — there is also a need to find its place in the order of things — some adding new layers to our understanding,

and others filling in gaps [2]. Classification systems are designed to manage the increasing volume of knowledge, so that it can be readily recorded, searched, explored and exploited in the pursuit to create new knowledge. From the Decimal Classification system (developed by expert committees and used in libraries around the world) to the Wikipedia category structure (a crowdsourced representation of our collective knowledge [85–88]), we are surrounded by ontologies that attempt to organize our common understanding. Likewise, the MeSH system aims to organize biomedical knowledge into a networked hierarchy that relates objects, methods, theory and other contextual metadata. Notably, the PACS system — a relatively shallow category system used to classify physics research — has recently been revamped into PhySH [89] by adopting a similar concept-based hierarchy better suited for capturing the multiple sub-disciplines and complex relations of theory and experiment that defining physics, which provides additional practical evidence for the value of high-resolution ontologies.

Here we utilized the MeSH ontology to analyze the entire PubMed index over a roughly half-century period, thereby contributing to the literature on convergence science, recombinant innovation, and the development of scientometric methods for analyzing interdisciplinary research. To address RQ1, we exploited the MeSH ontology structure to develop a dynamic and high-resolution map of convergence, operationalized by quantifying boundary-spanning cross-domain integration [18, 23, 24], which contributes to prior efforts to visualize the biomedical knowledge network [2, 9, 27, 28]. In particular, we developed two complementary measures of cognitive distance between knowledge entities, yielding visualization of MeSH co-occurrence on the one hand, and quantification of article-level cross-domain diversity on the other hand. Cross-temporal analysis indicates that the robust increasing trend in convergence $f_D(t)$ shown in Fig. 6(c), which features a 40% increase over the entire study period from 1970 to 2018. This increase is partly attributable to the decreasing prevalence of mono-domain research ($f_{D,p} = 0$) illustrated in Fig. 6(d), which is the analog to the drastically reduced frequency of solo-authored research observed over roughly the same period [48].

Structural analysis of MeSH co-occurrence networks visualized in Fig. 2 identify three robust macro-knowledge clusters: (a) the vast universe of microscopic biological entities and structures; (b) systems, disease and diagnostics; and (c) biological and social phenomena capturing emergent properties, processes and functions. To identify periods of disruption within this conceptual model of science, we also developed a method for quantifying cross-temporal transitions in the macro-knowledge clusters by tracking individual MeSH as they fluctuate between clusters, as illustrated in Fig. 3. Research at the health, behavioral and brain science frontiers typically integrate multiple distinct knowledge domains, signaling the emergence and future potential of convergence science [18, 21–24, 29–33]. The convergence nexus of Health Care (N), Behavior and Behavior Mechanisms (F01) and Information Science (L01) is a prime example, making way for transdisciplinary brain science [24] to map

and model brain circuits [90] that are fundamental to addressing the grand challenge underlying the "global burden of mental disorders" [91, 92].

To address RQ2 we used the results of RQ1 to develop a cross-cluster bridge metric for identifying which MeSH facilitate domain-spanning knowledge creation. Our results identified two particular knowledge domains — "Technology, Industry, and Agriculture" (MeSH J01) and "Information Science" (MeSH L01) that are prominent convergence bridges (see Fig. 4). Research featuring MeSH belonging to J01 and L01 benefit from the highly generalizable, scalable and codifiable characteristic of techno-informatic tools and algorithms that accelerate scientific discovery by facilitating high-throughput, measurement, data integration and analysis, which are critical features of convergence science at the frontiers of brain, behavior and health science [21, 24, 30, 33, 69, 92]. Indeed, the same informatics tools have permeated other boundary-spanning convergence zones [21], such as computational social science [93] and the science of science [94, 95].

Further analysis of MeSH–MeSH co-occurrence networks in Fig. 5 addresses RQ3 by highlighting the role of techno-informatic convergence bridges in brain, behavior and health sciences. Figures 6(a) and 6(b) highlight the rapid emergence of the potent combination of techno-informatics (J + L) in tandem (measured by $f_X(t)$). This paradigm shift occurred in the early 1990s, coinciding with the onset of the genomics revolution wherein cross-disciplinary collaboration between scholars from traditional biology and computer science departments provides an early example of successful convergence. Indeed, research integrating SAs that span relatively larger disciplinary distances is more impactful when executed by cross-disciplinary teams as opposed to mono-disciplinary teams [24]. In addition to cross-disciplinary collaboration, another success factor during the genomics revolution was the consortium science funding model, whereby teams of teams organized with a common goal to share benefits equitably within and beyond institutional boundaries [23]. These results call for a better mechanistic understanding of the role that cross-disciplinary scholars play in mediating the integration of peripheral and core concepts in the knowledge network that overlays collaboration networks [18, 47, 96].

Cross-temporal analysis of MeSH convergence — operationalized as cross-domain knowledge diversity calculated at the article level, given by $f_D(t)$ — reveals a steady rise of cross-domain integration over the last half century, which complements the steady emergence of team science [48, 97–99]. In particular, Figs. 6(e) and 6(f) show how medium- and large-scale teams have a notable advantage integrating multiple research domains. Yet viewed from a different bureaucratic perspective, we observe variable levels and growth rates of convergence in journals, possibly owing to the multi-disciplinary composition (or lack thereof) of journal editorial boards — see Figs. 6(g) and 6(h).

In summary, techno-informatic capabilities are increasingly essential to progress at the biomedical frontier. As illustrated by Fig. 6(a), these distinct inputs are increasingly found as integrated complements — for the period 2000–2018 we estimate that $\langle f_\times \rangle \approx 8.2\%$ of research articles relied on techno-informatic capabilities in

tandem (J + L); this frequency corresponds to a 291% growth over the J + L convergence levels in 1980. If the future relies on harnessing the combined power of human and machine intelligence, then the development of *human-in-the-loop* Man-Machine Systems (J01.897.441) — "in which the functions of the man and the machine are interrelated and necessary for the operation of the system" — will require transdisciplinary domains such as bio-mechatronics to flourish by harnessing convergence [100, 101]. For this reason, scientists are increasingly in need of knowledge maps to navigate the realm of possibilities and to thereby which conceptual bridges to cross.

Acknowledgments

IP acknowledges funding from the Eckhard-Pfeiffer Distinguished Professorship Fund; IP and AMP acknowledge support from NSF grant 1738163 entitled "From Genomics to Brain Science". AMP acknowledges financial support from a Hellman Fellow award that was critical to completing this project. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies. DY and AMP contributed equally to this work.

Appendix A. Methodological Appendices

A.1. *Major MeSH descriptors: Refinement to first and second branch-level representation*

We used the 2020 MeSH classification tree, which includes 29,638 descriptors that are organized in a tree-like structure denoting their hierarchical relations. MeSH descriptors are assigned to biomedical publications through an indexing process performed by examiners at the US NLM. With on average 12 MeSH per article, this ontology facilitates topic mapping and topic co-occurrence analysis at multiple levels of specificity [28]. We restrict our analysis to the "Major Topic Headings" for each article, which are indicated in each PubMed article page by an asterisk * next to the MeSH term; these Major MeSH account for roughly 1 in 3 MeSH descriptors, and so with on average 4 Major MeSH these annotations are sufficient to identify the article's core content. As such, we use these publication-level MeSH to determine the topical SAs as indicated by the 10 pre-defined science-oriented MeSH branches (A, B, C, D, E, F, G, J, L, N).

For each publication p, we extracted the set of Major MeSH terms (as indicated by an asterisk in PubMed records) and represent them by the vector $\mathbf{m}_p$ containing 29,638 elements, each representing a distinct MeSH term. We then map each MeSH term to its corresponding tree location, which specifies its classification among the 10 science-oriented MeSH branches. To be specific, we define a generic operator O_M which takes the vector $\mathbf{m}_p$ and maps these counts to a combined count vector $\mathbf{SA}_p^{(2)}$

with 104 elements representing all L_2 MeSH: $O_M^{(2)}(\mathbf{m}_p) = \mathbf{SA}_p^{(2)}$. Similarly, we define the operator $O_M^{(1)}(\mathbf{SA}_p^{(2)}) = \mathbf{SA}_p^{(1)}$, which maps the L_2 MeSH counts to their L_1 branches.

In order to count co-occurrences, we take each $\mathbf{SA}$ and denote the binary SA vector $B_p = \texttt{Sign}(\mathbf{SA})$, which reduces each element value to either 0 or 1. For an article with M_p nonzero elements, we then count all $\binom{M_p}{2}$ pairwise permutations which we record in a normalized co-occurrence matrix, given by $\mathbf{M}_p^{(1)}$ ($\mathbf{M}_p^{(2)}$) for the L_1 (L_2) representation. Each matrix is normalized to unity such that the total of all subelements $\sum_{i,j}\mathbf{M}_{p,ij} = 1$. This normalization step ensures that each publication contributes an equal share to the annual co-occurrence matrix, given by $\mathbf{M}_y^{(1)} = \sum_{p\in y}\mathbf{M}_p^{(1)} = N(y)$, where $N(y)$ is the total number of articles analyzed from year y (with similar definition for the L_2 level).

A.2. *Calculation of the cross-cluster bridge score*

Motivated by the bridge centrality index [77], we define the knowledge bridge score β_i of node i, given by

$$\beta_i = \sum_{C_I, C_J \in C \text{ and } C_I \neq C_J} D_{IJ} W_{iJ}, \tag{A.1}$$

which first requires identifying nodes belonging to distinct clusters, calculated here using the Louvain algorithm [68]. Here C represents the set of connected clusters in the giant component of an undirected network, and C_J represents a cluster that is different from the cluster C_I containing a given node i. We define D_{IJ} to be the distance between cluster C_I and C_J, measured as the inverse of the sum of weights of edges between the clusters, represented as $D_{IJ} = W_{IJ}^{-1} = (\sum_{i\in C_I \text{ and } j\in C_J} W_{ij})^{-1}$. Similarly, W_{iJ} is the sum of weights of edges between node i and all other nodes belonging to cluster C_J. For our purposes, we define the link weights as empirical co-occurrence values, $W \equiv \mathbf{M}_t^{(2)}$. Hence, individual MeSH terms (nodes) that play key roles in bridging knowledge clusters have high β_i values; contrariwise, nodes that are only connected to nodes belonging to their own cluster have $\beta = 0$.

Since $\mathbf{M}_t^{(2)}$ tallies are proportional to the total number $N(t)$ of articles published in year t, which are generally increasing, we instead focus on the rank associated with each β_i score, denoted by $r_{\beta,i}$. To address how to compare ranks of nodes from clusters of varying size (measured as the total number of nodes within the cluster I, denoted by $|C_I|$), then we define the normalized Bridge rank score as $R_i = r_{\beta,i}/|C_I|$, as plotted in Fig. 4.

Based upon each time series $R_{i,t}$ calculated at the 1-year time resolution, we identified emerging bridges using the following criteria: (i) The node is on average ranked in the top-20 of its own cluster; (ii) the time series is at least half as long as the

entire observation period from 1970 to 2018; (ii) There is a significant positive or negative trend, as determined by linear regression, such that the trend coefficient P-value is smaller than 0.01; (iv) the trend coefficient magnitude is sufficiently large, in magnitude > 0.1. These criteria identify 8 emerging knowledge bridges: E05, F01, F02, G07, J01, L01, N04 and N05.

A.3. *Outer-product method for measuring categorical diversity and disparity*

We leverage a generic tensor-product method that takes as input a weighted vector and yields a scalar diversity measure based upon categorical mixing. While the result of our approach is nearly identical to the Blau index (also referred to as the Gini–Simpson index), our diversity measure is motivated by way of dyadic co-occurrence rather than the standard formulation motivated around repeated sampling. Within the present context, the weighted vector of category counts is the SA decomposition of an article, given by $\mathbf{SA}_p^{(1)}$ or $\mathbf{SA}_p^{(2)}$. The resulting metric, represented by $f_{D,p}$, quantifies the degree of cross-domain co-occurrence and is a bounded statistic in the range $[0, 1)$.

Calculating $f_{D,p}$ begins with the outer-product $\mathbf{SA}_p \otimes \mathbf{SA}_p$, where $\otimes$ is the outer tensor product. The resulting matrix represents dyadic combinations of categories as opposed to permutations (i.e. capturing the subtle difference between an undirected and directed network). While we did not explore it further, this matrix formulation may also give rise to higher-order measures of diversity associated with the eigenvalues of the outer-product matrix.

We normalize the resulting outer-product matrix so that our analysis is not systematically biased by increase in the number of MeSH per paper over time; see [28] for secular growth in the number of MeSH per article, reflecting growth of the MeSH ontology accompanied by growth in the length and breadth of published research. All steps together, the normalized co-occurrence matrix $\mathbf{D}_p$ is given by

$$\mathbf{D}_p \equiv \frac{U(\mathbf{SA}_p \otimes \mathbf{SA}_p)}{\|U(\mathbf{SA}_p \otimes \mathbf{SA}_p)\|}. \tag{A.2}$$

Without loss of generality, this definition involves $U(\mathbf{G})$, an operator yielding the upper-diagonal elements of the arbitrary matrix $\mathbf{G}$ (i.e. representing the undirected co-occurrence network among subcategories); $\|\cdots\|$ indicates the matrix total calculated by summing across all matrix elements. The subtle difference between the Blau index arises from $U(\mathbf{G})$, which is imposed to capture the difference between combinations rather than permutations (or directed versus undirected network). Hence, this perspective offers a new pathway to this fundamental diversity measure by way of co-occurrence rather than repeated sampling.

Hence, $\mathbf{D}_p(\mathbf{SA}_p)$ tabulates the weighted product across all undirected SA–SA pairs. We then define the co-occurrence diversity as the scalar quantity given by

$$f_{D,p} = 1 - \mathrm{Tr}(\mathbf{D}_p) \in [0, 1), \tag{A.3}$$

where Tr is the matrix trace, corresponding to the sum of diagonal elements in $\mathbf{D}_p$. Hence, $f_{D,p}$ is the tally of off-diagonal elements in $\mathbf{D}$. Articles featuring a single SA have value $f_{D,p} = 0$, whereas articles featuring multiple SA have values in the range $0 < f_{D,p} < 1$. Regarding the upper limit, when all vector elements have equal values then $f_{D,p} = (d-1)/(d+1)$, where d is the dimension of the categorical vector (for the Blau index the upper limit is instead $d - 1/d$). In the present case $d = 10$ and so the maximum $f_{D,p}$ value is $9/11$. The average article diversity by publication year, denoted by $\langle f_D(t) \rangle$, is representative of a characteristic article since $f_{D,p} \in [0, 1)$ is bounded.

By way of example, consider a decomposition across only 6 SA, and the particular case of an article with 4 MeSH belonging to 3 different SA, e.g. $\mathbf{SA}_p = \{1, 2, 0, 0, 1, 0\}$. Calculation of the co-occurrence matrix $\mathbf{D}_p(\mathbf{SA}_p)$ in Eq. (A.2) yields

$$\mathbf{D}_p(\mathbf{SA}_p) = \begin{pmatrix} \frac{1}{11} & \frac{2}{11} & 0 & 0 & \frac{1}{11} & 0 \\ 0 & \frac{4}{11} & 0 & 0 & \frac{2}{11} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{11} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

with $\|U(\mathbf{SA}_p \otimes \mathbf{SA})\| = 11$. The categorical diversity is the total across the off-diagonal elements, $f_{D,p} = 5/11$.

For comparison, consider the representation of an article with the same number of metadata entities that all fall into just the second category, $\mathbf{SA}_p = \{0, 4, 0, 0, 0, 0\}$. In which case

$$\mathbf{D}_p(\mathbf{SA}_p) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

and so $f_{D,p} = 1 - 1 = 0$.

What does this measure measure? Notably, $f_{D,p}$ accounts for both categorical differences (Shannon-like) and concentration disparity (Gini-like) [80]. One the first hand, articles with more variation in SA categories will correspond to larger $f_{D,p}$

values, as the number of nonzero off-diagonal elements is proportional to $\binom{M_p}{2} \sim M_p^2$, where M_p is the number of distinct SA present, which contributes to larger $f_{D,p}$; and on the second hand, the off-diagonal elements will be relatively larger in combination if the count values contained in SA_2 are more evenly distributed, i.e. are not highly concentrated in just one category.

References

[1] Fleming, L., Recombinant uncertainty in technological search, *Manage. Sci.* **47** (2001) 117–132.
[2] Petersen, A. M., Evolution of recombinant biomedical innovation quantified via billions of distinct article-level MeSH keyword combinations, *Adv. Complex Syst.* **24** (2022) 2150016.
[3] Börner, K., *Atlas of Science: Visualizing What We Know* (MIT Press, 2010).
[4] Borner, K., *Atlas of Knowledge: Anyone Can Map* (MIT Press, 2015).
[5] Börner, K., *Atlas of Forecasts: Modeling and Mapping Desirable Futures* (MIT Press, 2021).
[6] Börner, K., Chen, C. and Boyack, K. W., Visualizing knowledge domains, *Annu. Rev. Inf. Sci. Technol.* **37** (2003) 179–255.
[7] Fleming, L. and Sorenson, O., Science as a map in technological search, *Strateg. Manage. J.* **25** (2004) 909–928.
[8] Börner, K., Klavans, R., Patek, M., Zoss, A. M., Biberstine, J. R., Light, R. P., Larivière, V. and Boyack, K. W., Design and update of a classification system: The UCSD map of science, *PLoS One* **7** (2012) e39464.
[9] Shi, F., Foster, J. G. and Evans, J. A., Weaving the fabric of science: Dynamic network models of science's unfolding structure, *Social Networks* **43** (2015) 73–85.
[10] Saket, B., Scheidegger, C., Kobourov, S. G. and Börner, K., Map-based visualizations increase recall accuracy of data, in *Computer Graphics Forum*, Vol. 34 (Wiley Online Library, 2015), pp. 441–450.
[11] Barry, A., Born, G. and Weszkalnys, G., Logics of interdisciplinarity, *Econ. Soc.* **37** (2008) 20–49.
[12] Cao, K.-A. L., González, I. and Déjean, S., Integromics: An R package to unravel relationships between two omics datasets, *Bioinform.* **25** (2009) 2855–2856.
[13] Gomez-Cabrero, D., Abugessaisa, I., Maier, D., Teschendorff, A., Merkenschlager, M., Gisel, A., Ballestar, E., Bongcam-Rudloff, E., Conesa, A. and Tegnér, J., Data integration in the era of omics: Current and future challenges, *BMC Syst. Biol.* **8** (2014) 1–10.
[14] Pan, R. K., Petersen, A. M., Pammolli, F. and Fortunato, S., The memory of science: Inflation, myopia, and the knowledge network, *J. Informetr.* **12** (2018) 656–678.
[15] Petersen, A. M., Pan, R. K., Pammolli, F. and Fortunato, S., Methods to account for citation inflation in research evaluation, *Res. Policy* **48** (2018) 1855–1865.
[16] Rzhetsky, A., Foster, J. G., Foster, I. T. and Evans, J. A., Choosing experiments to accelerate collective discovery, *Proc. Natl. Acad. Sci.* **112** (2015) 14569–14574.
[17] Helbing, D., Accelerating scientific discovery by formulating grand scientific challenges, *Eur. Phys. J. Spec. Top.* **214** (2012) 41–48.
[18] Petersen, A. M., Arroyave, F. and Pavlidis, I., Methods for measuring social and conceptual dimensions of convergence science, SSRN e-print: 4117933 (2022), pp. 1–11.

[19] Kuhn, T. S., The essential tension: Tradition and innovation in scientific research, in *The Third University of Utah Research Conference on the Identification of Scientific Talent* (University of Utah Press, Salt Lake City, 1959), pp. 162–174.
[20] March, J. G., Exploration and exploitation in organizational learning, *Organ. Sci.* **2** (1991) 71–87.
[21] Roco, M., Bainbridge, W., Tonn, B. and Whitesides, G., *Converging Knowledge, Technology, and Society: Beyond Convergence of Nano-Bio-Info-Cognitive Technologies* (Springer, New York, 2013).
[22] Pavlidis, I., Akleman, E. and Petersen, A. M., From polymaths to Cyborgs – Convergence is relentless, *Am. Sci.* **110** (2022) 196–200.
[23] Petersen, A. M., Majeti, D., Kwon, K., Ahmed, M. E. and Pavlidis, I., Cross-disciplinary evolution of the genomics revolution, *Sci. Adv.* **4** (2018) eaat4211.
[24] Petersen, A. M., Ahmed, M. E. and Pavlidis, I., Grand challenges and emergent modes of convergence science, *Humanit. Soc. Sci. Commun.* **8** (2021) 194.
[25] Shu, F., Julien, C.-A., Zhang, L., Qiu, J., Zhang, J. and Larivière, V., Comparing journal and paper level classifications of science, *J. Informetr.* **13** (2019) 202–225.
[26] MeSH, The MeSH (Medical Subject Headings) system: A controlled vocabulary thesaurus used for indexing PubMed articles (2020), http://www.ncbi.nlm.nih.gov/mesh.
[27] Leydesdorff, L., Rotolo, D. and Rafols, I., Bibliometric perspectives on medical innovation using the medical subject headings of pub med, *J. Am. Soc. Inf. Sci. Technol.* **63** (2012) 2239–2253.
[28] Petersen, A. M., Rotolo, D. and Leydesdorff, L., A triple helix model of medical innovation: Supply, demand, and technological capabilities in terms of medical subject headings, *Res. Policy* **45** (2016) 666–681.
[29] National Research Council, *Facilitating Interdisciplinary Research* (National Academies Press, Washington, D.C., 2005).
[30] Sharp, P. A. and Langer, R., Promoting convergence in biomedical science, *Science* **333** (2011) 527.
[31] National Research Council, *Convergence: Facilitating Transdisciplinary Integration of Life Sciences, Physical Sciences, Engineering, and Beyond* (National Academies Press, Washington, D.C., 2014).
[32] Linkov, I., Wood, M. and Bates, M., Scientific convergence: Dealing with the elephant in the room, *Environ. Sci. Technol.* **48** (2014) 10539–10540.
[33] Eyre, H. A., Lavretsky, H., Forbes, M., Raji, C., Small, G., McGorry, P., Baune B. T. and Reynolds, C., Convergence science arrives: How does it relate to psychiatry?, *Acad. Psychiatry* **41** (2017) 91–99.
[34] Lipsey, R. G., Carlaw, K. I. and Bekar, C. T., *Economic Transformations: General Purpose Technologies and Long-Term Economic Growth* (OUP, Oxford, 2005).
[35] Bonaccorsi, A., Search regimes and the industrial dynamics of science, *Minerva* **46** (2008) 285.
[36] Mariotti, F. and Haider, S., Managing institutional diversity and structural holes: Network configurations for recombinant innovation, *Technol. Forecast. Soc. Change* **160** (2020) 120237.
[37] Pyka, A. and Scharnhorst, A., *Innovation Networks: New Approaches in Modelling and Analyzing* (Springer Science & Business Media, 2010).
[38] Acemoglu, D., Akcigit, U. and Kerr, W. R., Innovation network, *Proc. Natl. Acad. Sci.* **113** (2016) 11483–11488.
[39] Leydesdorff, L. and Etzkowitz, H., Emergence of a triple helix of university–industry–government relations, *Sci. Public Policy* **23** (1996) 279–286.

[40] Etzkowitz, H. and Leydesdorff, L., The dynamics of innovation: From national systems and "Mode 2" to a triple helix of university–industry–government relations, *Res. Policy* **29** (2000) 109–123.
[41] Bromham, L., Dinnage, R. and Hua, X., Interdisciplinary research has consistently lower funding success, *Nature* **534** (2016) 684–687.
[42] NSF, NSF convergence accelerator (2019), https://www.nsf.gov/od/oia/convergence-accelerator/.
[43] Colón, W., Chitnis, P., Collins, J. P., Hicks, J., Chan, T. and Tornow, J. S., Chemical biology at the us national science foundation, *Nat. Chem. Biol.* **4** (2008) 511–514.
[44] Pan, R. K., Sinha, S., Kaski, K. and Saramäki, J., The evolution of interdisciplinarity in physics research, *Sci. Rep.* **2** (2012) 1–8.
[45] Leahey, E. and Moody, J., Sociological innovation through subfield integration, *Soc. Curr.* **1** (2014) 228–256.
[46] Helbing, D., Globally networked risks and how to respond, *Nature* **497** (2013) 51–59.
[47] Fleming, L., Perfecting cross-pollination, *Harv. Bus. Rev.* **82** (2004) 22–24.
[48] Wuchty, S., Jones, B. F. and Uzzi, B., The increasing dominance of teams in production of knowledge, *Science* **316** (2007) 1036–1039.
[49] Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., Rafols, I. and Börner, K., Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature, *J. Informetr.* **5** (2011) 14–26.
[50] Yegros-Yegros, A., Rafols, I. and Deste, P., Does interdisciplinary research lead to higher citation impact? The different effect of proximal and distal interdisciplinarity, *PLoS One* **10** (2015) e0135095.
[51] Porter, A., Cohen, A., Roessner, J. D. and Perreault, M., Measuring researcher interdisciplinarity, *Scientometrics* **72** (2007) 117–147.
[52] Porter, A. and Rafols, I., Is science becoming more interdisciplinary? Measuring and mapping six research fields over time, *Scientometrics* **81** (2009) 719–745.
[53] Rotolo, D. and Petruzzelli, A. M., When does centrality matter? Scientific productivity and the moderating role of research specialization and cross-community ties, *J. Organ. Behav.* **34** (2013) 648–670.
[54] Boyack, K. W., Mapping knowledge domains: Characterizing PNAS, *Proc. Natl. Acad. Sci.* **101** (2004) 5192–5199.
[55] Petersen, A. M., Megajournal mismanagement: Manuscript decision bias and anomalous editor activity at PLoS One, *J. Informetr.* **13** (2019) 100974.
[56] Fleming, L., Breakthroughs and the "long tail" of innovation, *MIT Sloan Manage. Rev.* **49** (2007) 69.
[57] Youn, H., Strumsky, D., Bettencourt, L. M. and Lobo, J., Invention as a combinatorial process: Evidence from us patents, *J. R. Soc. Interface* **12** (2015) 20150272.
[58] Verhoeven, D., Bakker, J. and Veugelers, R., Measuring technological novelty with patent-based indicators, *Res. Policy* **45** (2016) 707–723.
[59] Boyack, K. W., Newman, D., Duhon, R. J., Klavans, R., Patek, M., Biberstine, J. R., Schijvenaars, B., Skupin, A., Ma, N. and Börner, K., Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches, *PLoS One* **6** (2011) e18029.
[60] Foster, J. G., Rzhetsky, A. and Evans, J. A., Tradition and innovation in scientists' research strategies, *Am. Sociol. Rev.* **80** (2015) 875–908.
[61] Mane, K. K. and Börner, K., Mapping topics and topic bursts in PNAS, *Proc. Natl. Acad. Sci.* **101** (2004) 5287–5290.
[62] Palchykov, V., Krasnytska, M., Mryglod, O. and Holovatch, Y., Network of scientific concepts: Empirical analysis and modeling, *Adv. Complex Syst.* **24** (2021) 214001.

[63] Nissani, M., Fruits, salads, and smoothies: A working definition of interdisciplinarity, *J. Educ. Thought* **29** (1995) 119–126.
[64] Pedersen, D. B., Integrating social sciences and humanities in interdisciplinary research, *Palgrave Commun.* **2** (2016) 1–7.
[65] Arroyave, F. J., Goyeneche, O. Y., Gore, M., Heimeriks, G., Jenkins, J. and Petersen, A. M., On the social and cognitive dimensions of wicked environmental problems characterized by conceptual and solution uncertainty, *Adv. Complex Syst.* **24** (2021) 2150005.
[66] Kavuluru, R. and Rios, A., Automatic assignment of non-leaf mesh terms to biomedical articles, in *AMIA Annual Symp. Proc.*, Vol. 2015 (American Medical Informatics Association, 2015), p. 697.
[67] MeSH Tutorial, MeSH Changes and PubMed Searching (2018), https://www.youtube.com/watch?v=K-lN2SAm2KQ, Contributed: 2018-02-07.
[68] Blondel, V. D., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E., Fast unfolding of communities in large networks, *J. Stat. Mech., Theory Exp.* **2008** (2008) P10008.
[69] Sadybekov, A. V. and Katritch, V., Computational approaches streamlining drug discovery, *Nature* **616** (2023) 673–685.
[70] Doudna, J. A. and Charpentier, E., The new frontier of genome engineering with CRISPR-Cas9, *Science* **346** (2014) 1258096.
[71] Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S., Induction of pluripotent stem cells from adult human fibroblasts by defined factors, *Cell* **131** (2007) 861–872.
[72] Benner, S. A. and Sismour, A. M., Synthetic biology, *Nat. Rev. Genet.* **6** (2005) 533–543.
[73] Church, G. M. and Regis, E., *Regenesis: How Synthetic Biology Will Reinvent Nature and Ourselves* (Basic Books, 2014).
[74] Hoshika, S. *et al.*, Hachimoji DNA and RNA: A genetic system with eight building blocks, *Science* **363** (2019) 884–887.
[75] Fleming, L. and Frenken, K., The evolution of inventor networks in the silicon valley and boston regions, *Adv. Complex Syst.* **10** (2007) 53–71.
[76] Tomasello, M. V., Tessone, C. J. and Schweitzer, F., A model of dynamic rewiring and knowledge exchange in R&D networks, *Adv. Complex Syst.* **19** (2016) 1650004.
[77] Jensen, P., Morini, M., Karsai, M., Venturini, T., Vespignani, A., Jacomy, M., Cointet, J.-P., Mercklé, P. and Fleury, E., Detecting global bridges in networks, *J. Complex Networks* **4** (2016) 319–329.
[78] Markram, H., The human brain project, *Sci. Am.* **306** (2012) 50–55.
[79] Grillner, S., Ip, N., Koch, C., Koroshetz, W., Okano, H., Polachek, M., Poo, M.-M. and Sejnowski, T. J., Worldwide initiatives to advance brain research, *Nat. Neurosci.* **19** (2016) 1118–1122.
[80] Harrison, D. A. and Klein, K. J., What's the difference? Diversity constructs as separation, variety, or disparity in organizations, *Acad. Manage. Rev.* **32** (2007) 1199–1228.
[81] Heller, A., *Renaissance Man* (Routledge, 2015).
[82] Simonton, D. K., After Einstein: Scientific genius is extinct, *Nature* **493** (2013) 602.
[83] Mryglod, O., Kenna, R., Holovatch, Y. and Berche, B., Absolute and specific measures of research group excellence, *Scientometrics* **95** (2013) 115–127.
[84] Kenna, R. and Berche, B., Critical masses for academic research groups and consequences for higher education research policy and management, *High. Educ. Manage. Policy* **23** (2012) 1–21.
[85] Suchecki, K., Salah, A. A. A., Gao, C. and Scharnhorst, A., Evolution of wikipedia's category structure, *Adv. Complex Syst.* **15** (2012) 1250068.

[86] Salah, A. A., Gao, C., Suchecki, K. and Scharnhorst, A., The need to categorize: A comparative look at categorization in wikipedia and the universal decimal classification system, *Leonardo* **45** (2012) 84–85.

[87] Salah, A. A. A., Gao, C., Scharnhorst, A. and Suchecki, K., Design vs. emergence: Visualisation of knowledge orders, *Places & Spaces: Mapping Science: 7th Iteration (2011): Science Maps as Visual Interfaces to Digital Libraries.*

[88] Scharnhorst, A., Smiraglia, R. P., Guéret, C. and Salah, A. A. A., Knowledge maps of the UDC: Uses and use cases, *Knowl. Organ.* **43** (2016) 641–654.

[89] Smith, A., From PACS to PhySH, *Nat. Rev. Phys.* **1** (2019) 8–11.

[90] Jorgenson, L. A. *et al.*, The brain initiative: Developing technology to catalyse neuroscience discovery, *Philos. Trans. R. Soc. B: Biol. Sci.* **370** (2015) 20140164.

[91] Kessler, R. C., Aguilar-Gaxiola, S., Alonso, J., Chatterji, S., Lee, S., Ormel, J., Üstün, T. B. and Wang, P. S., The global burden of mental disorders: An update from the WHO World Mental Health (WMH) surveys, *Epidemiol. Psychiatr. Sci.* **18** (2009) 23–33.

[92] Dzau, V. J. and Balatbat, C. A., Reimagining population health as convergence science, *Lancet* **392** (2018) 367–368.

[93] Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D. and Alstyne, M. V., Life in the network: The coming age of computational social science, *Science* **323** (2009) 721–723.

[94] Fortunato, S., Bergsrom, C. T., Borner, K., Evans, J. A., Helbing, D., Milojevic, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D. and Barabasi, A.-L., Science of science, *Science* **359** (2018) eaao0185.

[95] Verginer, L., Vaccario, G. and Petersen, A. M., Foreword to the special issue on success in science, *Adv. Complex Syst.* **24** (2021) 2102001.

[96] Schweitzer, F., Garas, A., Tomasello, M. V., Vaccario, G. and Verginer, L., The role of network embeddedness on the selection of collaboration partners: An agent-based model with empirical validation, *Adv. Complex Syst.* **25** (2022) 2250003.

[97] Milojevic, S., Principles of scientific research team formation and evolution, *Proc. Natl. Acad. Sci.* **111** (2014) 3984–3989.

[98] Pavlidis, I., Petersen, A. M. and Semendeferi, I., Together we stand, *Nat. Phys.* **10** (2014) 700.

[99] Petersen, A. M., Pavlidis, I. and Semendeferi, I., A quantitative perspective on ethics in large team science, *Sci. Eng. Ethics* **20** (2014) 923–945.

[100] Li, Z., Yang, C. and Burdet, E., Guest editorial an overview of biomedical robotics and bio-mechatronics systems and applications, *IEEE Trans. Syst. Man Cybern: Syst.* **46** (2016) 869–874.

[101] Kose, T. and Sakata, I., Identifying technology convergence in the field of robotics research, *Technol. Forecast. Soc. Change* **146** (2019) 751–766.