THE UNIVERSITY OF CHICAGO

Where do new ideas come from: New directions in science emerge from disconnection and discord

By

Yiling Lin

August 2021

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts In Computational Social Science

Faculty Advisor: James Allen Evans
Preceptor: Pedro Alberto Arroyo

***This work is collaborating with Prof. James Evans and Prof. Lingfei Wu (University of Pittsburgh). The original paper can be found at: https://arxiv.org/abs/2103.03398**

Abstract

Since the 1950s, citation number or "impact" has been the dominant metric by which science is quantitatively evaluated. But research contributions play distinct roles in the unfolding drama of scientific debate, agreement and advance, which are differentially valued by scientists and their institutions. Computational power, access to citation data and an array of modeling techniques have given rise to a widening portfolio of metrics that extract different signals regarding their contribution to scientific activities. Besides impact, the innovation of work and the way that it builds up the scientific discussions are taken into consideration. Here we unpacks the complex, temporally evolving relationship between citation impact alongside novelty and disruption, two emerging measures that capture the degree to which science not only influences, but transforms later work. Novelty is measured at the point of production and captures how research draws upon unusual combinations of prior work. Disruption is measured over time and captures how research comes to eclipse or amplify the prior work on which it builds. We theorize that novel papers will exhibit disruptive impact over time, and demonstrate how they are much more likely than conventional papers to disrupt current literature. Novel papers do not do so immediately, but often become "sleeping beauties", accumulating surprising attention and citation impact over the long run. In summary, new directions for science are created from a lack of consensus. Finally, we show how novelty can be reformulated as the combination of ideas across knowledge space to reveal the combinatorial nature of advance. The evolution of knowledge space over time characterizes how yesterday's novelty forms today's scientific conventions, which condition the novelty—and surprise—of tomorrow's breakthroughs.

Section 1: Introduction

Bruno Latour depicts a beautiful connection between people and knowledge in his book--science in action: big inquiries are explored by people and people form networks to make science that they believe (Latour 1987). It is true no matter in today or in the past that science proceeds with new ideas and new bloods to nurture the ideas. New ideas do not come from ex nihilo. Dating back to the history of creation in art, technology and science, every generation of new ideas can find their origins in earlier times. Some part of Beethoven's music is developed from that of Mozart, the changing world technique-- touch screen of Apple also developed from hundreds of patents, and the discovery of double helix structure of DNA was supported by the theory of "Chargaff laws". Merton conceptualized this process of creation as a recombination of knowledge (Merton 1973).

Where do new ideas in science come from? How can a recombination be regarded as a novelty? There are several social processes associated with it. The first social process is labor division. In natural sciences like astronomy, diversified expertise is piled into a giant project to construct a best dataset. Labor division is characterized by the growing dominance of team cooperation in research activities (Wuchty, Jones, and Uzzi 2007). The second process is brokerage. Burt suggests that brokerage and closure are two complementary structures on social networks for creating the vision of good ideas (Burt 2007). Ideas are introduced from an established domain to a foreign domain and they make innovative impacts in that foreign one (J. Wang, Veugelers, and Stephan 2016). This process is marked by the cross-disciplines study. The third process is the formation of a new community. The boundary between disciplines is getting blurred and new directions emerge: for example, digital data together with social sciences reveals the complexity of human behaviors at different levels, signaling the new research opportunities and attracting brilliant minds from different domains (D. M. J. Lazer et al. 2020).

The various ways of combinations propose a question. The existing ideas are the building blocks of the future new ideas, and why are novel ideas still so sparse compared with the building blocks that are already accumulated by human creation? To answer this question, two concepts need to be considered. Ideas in science and novelty of them. Ideas are unscalable. Several good ideas that are combined together do not guarantee the new idea a better one. Big science like collecting astronomic data around the world by hundreds of scientists does not mean it is hundreds of times more important than work by a single author in the same field. And this insight is encapsulated into the concept: instrument (Latour 1987). Combination of ideas in science and technology are built into instruments and diffused within the community. For example, Bidirectional Encoder Representations from Transformers (BERT) encapsulates creative thoughts from semi-supervised sequence learning, generative pre-training and so on into a single power language model (Devlin et al. 2018). No matter how much knowledge is encapsulated together, the idea in the making is still considered as a single instrument once it is made (Latour 1987). Furthermore, the number of laborers in science may scale up the coverage of the topics and thus its impact to its community (B. Uzzi et al. 2013) but not necessarily the innovation it brings in. On the contrary, studies suggest that large teams are not as innovative in knowledge making as the small teams (Funk and Owen-Smith 2016; Wu, Wang, and Evans 2019). A more popular and welcomed instrument does not mean it is more scientific in its effect on the future of study.

Second, novelty does not equal the diversity of knowledge combined. Novelty in science is shaped by two forces in science: expanding to new ideas and validating the existing thoughts. The system of science is interwoven with enhancing consensus and bringing in new insight. The seeking of novelty This dichotomy reflects a fundamental tension identified by many scholars under different names: “conformity vs. dissent” (Polanyi 1962), “succession vs. subversion” (Bourdieu 1975), paradigm “deepening vs. changing” (Dosi 1982; Ahuja, Lampert, and Tandon 2014), “enhancing vs. destroying” (Tushman and Anderson 1986), “exploitation vs. exploration” (March 1991), “relevance vs. originality” (Whitley 2000), “conventionality vs. novelty” (Brian Uzzi et al. 2013), “tradition vs. innovation” (Foster, Rzhetsky, and Evans 2015), “destabilization vs. consolidation” (Chen, Shao, and Fan 2021), or path “deepening vs. breaking” (Garud, Kumaraswamy, and Karnøe 2010; Karim and Mitchell 2017). Recent empirical research on innovation in science defined novelty as the distance on the knowledge space. Empirically, there are qualitative and quantitative methods to define the distance. Novelty can be seen as successfully linking the unlinked concepts (Hofstra et al. 2020), and the atypicality of journals a paper combines (B. Uzzi et al. 2013; J. Wang, Veugelers, and Stephan 2016). It is suggested that new ideas should keep a balance between conventionality and novelty to make a major impact in science (B. Uzzi et al. 2013; Burt 2007).

Another reason for our perception that novel ideas in science are sparse is that visibility is a narrative concept that is used as a measurement to the impact. Commonly used measurement of impact is the number of citations. However, how to evaluate the significance of work is an open question. First, the quality of citation is doubted and Leiden manifesto suggests peer reviews is a more reliable way than the citation number since the peers are responsible for their reviews(Hicks et al. 2015). Second, the citation number depicts the tendency of research topics and is biased towards popular fields of study. Indexes other than citation are needed for evaluating the novel impact of new ideas. A recent, prominent metric has arisen that highlights work generating new combinations and directions, contributing to the “creative destruction” of science (McMahan and McFarland 2021). Disruption models how research comes to eclipse the prior work on which it builds, becoming recognized as a new scientific direction (Funk and Owen-Smith 2016; Wu, Wang, and Evans 2019). Disruption is the relative position of a paper in the citation network compared with its references. If the number of prospective works cite it instead of its references outweigh that of works that cite it together with one or more of its references, then this work is disruptive with D score >0. On the contrary, If the number of prospective works cite it and one or more of its references outweigh that of works that cite it instead of its references, then this work is developing with D score <0 (Funk and Owen-Smith 2016; Wu, Wang, and Evans 2019). D score can be understanded as expanding science vs. validating science. Based on this insight, Figure 1 depicts two prototypes of science: disrupting vs. developing where almost works within are disruptive or developing. The two prototypes model the battle between expanding and validating forces in science and our science in reality is a more developing one, as 67% of journal papers are developing with D score <0.

Where do new directions in science come from? Do we have better measurement for the new direction of science than being highly cited? In this work, I would like to explore how a work would be considered as a new direction in science and the structure of this work is listed as followed: section 2 hypothesize how a new direction for science is created; section 3 lists the data and methodology this work used and section 4 lists the findings and section 5 discuss the possible future directions on this question.

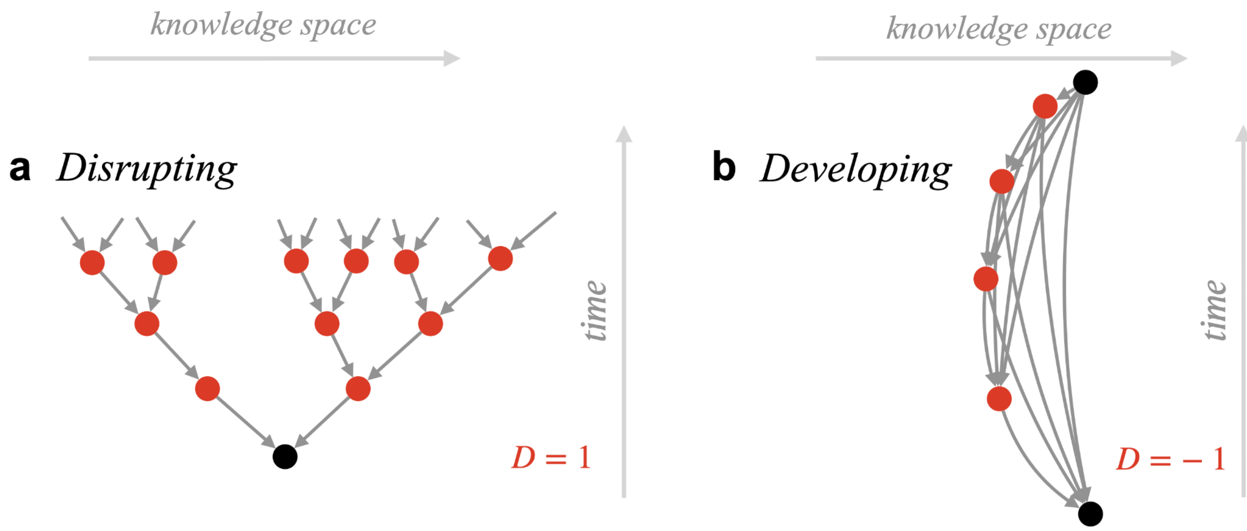

Figure 2. Illustration of two extreme citation structures of science. Science expands if the majority of papers (red circles) are disruptive (D = 1, Panel a) and collapse if developing (D = -1, Panel b). In the first scenario, science keeps branching out to cover new topics, but no consensus can be achieved. In the second scenario, science keeps returning back to the same topic and new papers are not significantly different from old papers after many generations—consensus is well established, but no new ideas are possible.

Section 2: Hypothesis

At first, we categorize the new ideas in the making based on how they combine the previous ideas, as shown in Figure 2.

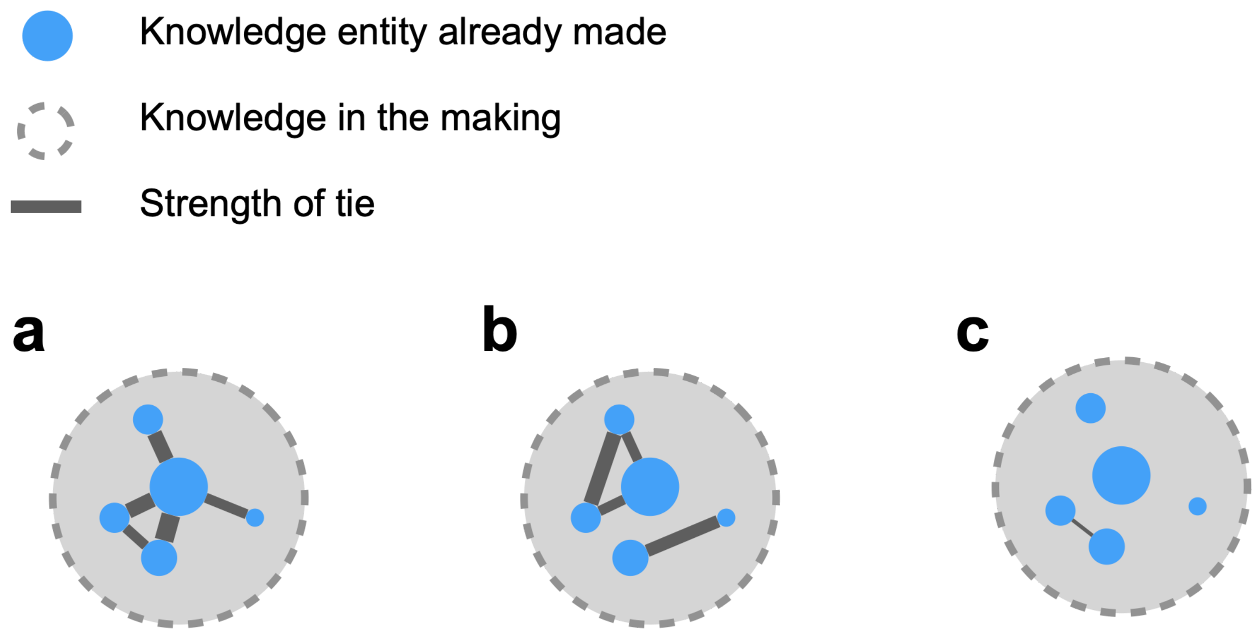

Figure 2. Three types of ideas in the making. New ideas may appear from established consensus in science (panel a) or bring established ideas into new fields (panel b) or from the lack of consensus (panel c).

Uzzi et al. suggests that works that bring in new thoughts into established communities have a higher chance to be highly cited (B. Uzzi et al. 2013). Under measurement of citation, works of type b (Fig. 2) are favored by the scientific community. However, the citation impact and D score depicts two different dynamics of diffusion of knowledge. The most striking difference is the time of receiving a citation. The citation time of a typical paper peaks at around three years after its publication (D. Wang, Song, and Barabási 2013), while disruptive work resembling a sleeping beauty is cited in a longer period of time. Time matters in the discussion of impact of new ideas. At first, turning points in science always take time to happen. The reason behind this is there is a bias towards novelty in science and it takes time to compensate, although not completely, for the bias. The risks of being novel are getting accepted by lower ranked journals, taking a longer time to be recognized and getting higher recognition in foreign fields than in their native fields (J. Wang, Veugelers, and Stephan 2016).

Science is a system seeking novelty and we would like to test how much novelty can this system accept. In this way, this work hypothesized that a new direction in science came from lack of consensus. As shown in Fig. 3.

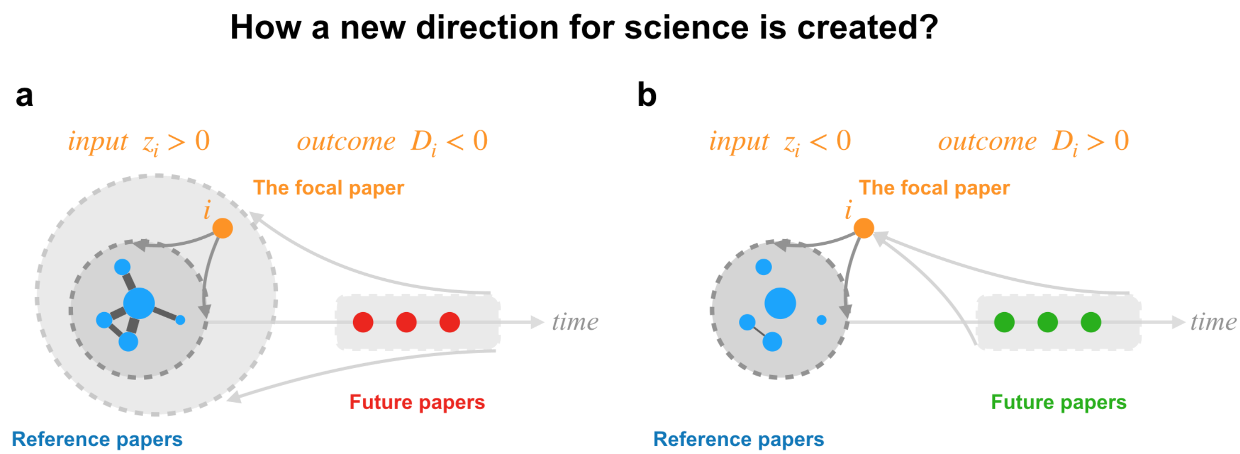

Figure 3. An illustration of how novelty and disruption are related. In Panel a, the focal paper draws upon and contributes to literature on a well studied topic, characterized by "clustered" references that have high, pairwise (Granovetter 1973). Network modularity (Newman 2006) emerges from these strong-ties, reflecting well-established consensus (Shwed and Bearman 2010). In Panel b, the focal paper identifies an unsolved problem and addresses it by integrating distant literature that is weakly connected or even disconnected ($z < 0$), which implies a lack of consensus. These two kinds of intellectual activities are not only different in input, but generate very different outcomes. The focal paper contributing to a developed topic is more likely to be viewed in future as a part of the ongoing conversation. For this reason, future papers (red circles in Panel a) will judge this paper as developing the broader topic or field (the larger gray circle in Panel a) and cite it together with its references ($D < 0$). On the contrary, the focal paper identifying unsolved problems tends to be viewed as creating a new direction in which future papers (green circles in Panel b) will cite it as a starting point, ignoring its references ($D > 0$).

Section 3: Data and Method

Data.

This work investigates impact, novelty and disruption using the Microsoft Academic Graph (MAG), which includes 87,860,684 journal articles published during 1800-2020 and 1,042,590,902 citations created by these articles. We calculated two variables for each of the 35,431,832 journal articles that have both citations and references, including D-score for disruption (higher is more disruptive) and z-score for novelty (lower is more novel). Our analysis of novelty, disruption, and citation impact covers multiple cohorts, including the cohort of 1970 (87,475 papers), 1980 (176,826 papers), 1990 (318,914 papers), and 2000 (591,653 papers).

These papers have 21.1 references on average. The median number of citations to these papers is 15.0. The construction of the journal embedding space is based on the co-citation of 7,774 journals by the papers published in 2000.

Methods.

Calculating the z-score of novelty. Uzzi et al. (2013) defined the z-score of a pair of journals co-cited in an article as more or less typical with z_{ij}:

$$z_{ij} = \frac{obs_{ij} - exp_{ij}}{\sigma_{ij}} \quad (1)$$

Where i and j are journals, obs_{ij} is the empirical frequency that these two journals co-referred across research articles (Brian Uzzi et al. 2013) and exp_{ij} is the expected frequency obtained through random shuffling. Random shuffling is implemented by repeatedly selecting two citation links at random with the same origin and target years and exchanging the papers to which they are attached. In this way, two variables remain unchanged, including the length of references for each paper and the temporal distribution characterizing references. After random shuffling, we recalculate exp_{ij}. We create ten random-shuffled versions to obtain the stabilized value of exp_{ij} and standard deviation σ_{ij} to normalize obs_{ij} into z_{ij}, following the practice suggested by (Brian Uzzi et al. 2013).

Calculating the D-score of disruption. The Disruption, *D*, of a focal paper, can be calculated as the difference between the fractions of two types of subsequent papers, type *i* that only cites the focal paper but not its reference and type *j* that cites both. A paper may disrupt earlier research by introducing new ideas that come to be recognized independent from the prior work on which it builds ($0 < D < 1$), develop existing research by providing supportive evidence or extension that come to be recognized as development on prior work ($-1 < D < 0$), or remain neutral, keeping in balance the disruptive and developmental character of its contribution ($D = 0$).

$$D = p_i - p_j = \frac{n_i - n_j}{n_i + n_j + n_k} \quad (2)$$

D-score may change over time due to the temporal evolution of the two kinds of subsequent papers. To calculate a stabilized disruption score, we used the longest time window available in the MAG dataset from the year of publication to 2018 for each paper under study. In section 2 of our findings, we explore the temporal dynamics of D(t), i.e., how D-score changes with time, for two field-definitive studies in biology, the paper on DNA by Watson and Crick (Watson and Crick 1953) and the paper on RNA by Baltimore (Baltimore 1970) and also four cohorts of papers published in 1970, 1980, 1990, and 2000.

Reformulating z-score novelty as distance in knowledge space

Z-score is deeply related to a common measure in information science, the Pointwise mutual information (*PMI*) between two items. These items may be represented by vectors, as constructed at scale within an embedding model like that produced by word2vec. We can derive a revised version of the z-score from PMI as

$$PMI_{ij} = log_2\left(\frac{P_{ij}}{P_i \times P_j}\right) = log_2(P_{ij}) - log_2(P_i \times P_j)$$
$$= log_2(obs_{ij}) - log_2(exp_{ij}) \qquad (3)$$

where P_i and P_j are the probabilities that *i* and *j* appear independently, respectively, and P_{ij} is the joint probability. The connection between PMI and z-score permits defining and measuring z-score as the distance on latent semantic spaces obtained through an embedding model, such as the popular skip-gram word2vec model, which has been demonstrated to preserve semantic compositionality within word vector sufficient to perform at human level on semantic analogy problems (*a* is to *b* as *c* is to ___?) (Mikolov et al. 2013). Word embedding models have inspired a wide range of item-context embedding models beyond words, ranging from images (Xian et al. 2016) and audio clips (Xie and Virtanen 2019) to graphs (Perozzi, Al-Rfou, and Skiena 2014; Grover and Leskovec 2016) and journals (Tshitoyan et al. 2019; Peng et al. 2020; Miao et al. 2021).

In an embedding model, each item is represented as a vector in shared vector space. For example, in a word embedding, words sharing similar contexts within the text will be positioned nearby in the space, whereas words that appear only in distinct and disconnected contexts will be positioned farther apart. The same holds for journals embedded as a function of their co-citation within reference lists. Consider the structure of the descriptive problem that embeddings attempt to solve: how to represent all items from a dataset within the *k*-dimensional space that best preserves distances between *n* items (e.g., journals) across *m* contexts (e.g., article reference lists). The solution, is a *n*-by-*k* matrix of values, where $k \ll m$. Early embedding approaches used singular-value decomposition (SVD) to factorize this item-context matrix, where contexts were large and nondiscriminating (e.g., entire documents of thousands or tens of thousands of words), but SVD placed strict upper limits on the number of contexts they could factorize. Neural embeddings use heuristic optimization of a neural network with at least one "hidden-layer" of *k* internal, dependent variables, enabling a factorization of much larger item-context matrices constructed from vast numbers of arbitrarily local item contexts (very large *m*).[1] In this way, PMI is formally equivalent to the inner product of two vectors representing items within a latent semantic space (Levy and Goldberg 2014). Specifically,

$$emb_{in-i} \cdot emb_{out-j} = PMI_{ij} - log_2 Neg \qquad (4)$$

where emb_{in-i} is the item embedding of i and emb_{out-j} is the context embedding of j. *Neg* is the number of negative samples per positive (actual item-context) sample. In sum, the inner product between journal vectors in an embedding space is a computationally efficient proxy for the z-score. In section 3 of findings, we will train journal vectors across time periods to visualize and compare the changing landscape of novelty in science.

Section 4: Findings

1. Novel papers are more likely to disrupt existing literature

Does a novel paper expand the scientific frontier, as represented in subsequent papers? Novelty captures how research draws upon unusual combinations of prior knowledge in crafting a new idea. It crystallizes at the moment of a works' creation. Disruption, by contrast, captures how research comes to eclipse the prior work on which it builds, becoming recognized as a new scientific direction. It evolves after creation and takes years to stabilize, when neither the focal

[1] Scientists have attempted to perform these parametrically, as with exponential family embedding models, but their performance has not yet approached that of autoencoders (Rudolph et al 2016).

work nor its references receive more citations. Novelty can be designed by scholars and prioritized by funding agencies, who change science through the projects they support, but disruption is hard to plan in advance. If novelty is associated with disruption, then science can be better managed and planned—the design and effort in experimenting with new ideas and testing bold assumptions will eventually pay off, and translate into new advances in science that support future applications.

To answer this question, we initially analyze the association between two measures of papers, z-score for novelty and D-score for disruption. Z-score measures how a focal paper integrates conventional or surprising journal pairs in its references (Brian Uzzi et al. 2013). Each pair of journals has a z-score, indicating whether they are frequently co-cited across the corpus ($z > 0$), such as "*American Sociological Review*" and "*American Journal of Sociology*", or not ($z < 0$), such as "*American Sociological Review*" and "*Physical Review Letters*" (see Methods for z-score calculation). Paper novelty is defined on the distribution of z-scores including all journal pairs across all reference lists. The novelty measure could be the 50th percentile of the z-score distribution, which captures median reference novelty, or the 10th percentile, which identifies extreme reference novelty. Z-score for reference citations can be understood as an analog of the concept of "tie strength" proposed by (Granovetter 1973). He described "strong ties" as redundant, repetitive links within communities and "weak ties" as bridges between them. An alternative version of z-score has been proposed using the analytical tools of statistical physics to identify network communities by grouping nodes connected by paths consisting of strong ties (Newman 2006). From this perspective, z-score is a useful indicator for distinguishing "structured" ($z > 0$) from "unstructured" references ($z < 0$).

Noting that z-score only quantifies content novelty at the moment of publication and does not consider the unfolding influence of a paper over time, we the "disruption" or D-score of a paper to capture recognized novelty and importance (Funk and Owen-Smith 2016; Wu, Wang, and Evans 2019). Disruption measures the extent to which a focal paper eclipses the contributions on which it builds. The disruption of a focal paper is defined as the difference between two fractions of subsequent papers, type *i* that only cite the focal paper but not its references, and type *j* that cite both, giving $D = pi - pj$. A paper may disrupt its prior work by introducing new ideas (0 <

$D < 1$), develop existing theories by providing supportive evidence ($-1 < D < 0$), or remain neutral ($D = 0$) (see Methods for D-score calculation).

We find that novelty (z-score) and disruption (D-score) are associated. Papers integrating unusual combinations of literature come to be seen as disruptive by a disproportionate number of subsequent papers that only cite those novel papers but not their references. In comparison, papers drawing upon typical combinations of references are deemed as developing prior approaches by the majority of following papers that cite those papers in context with their references—as extensions.

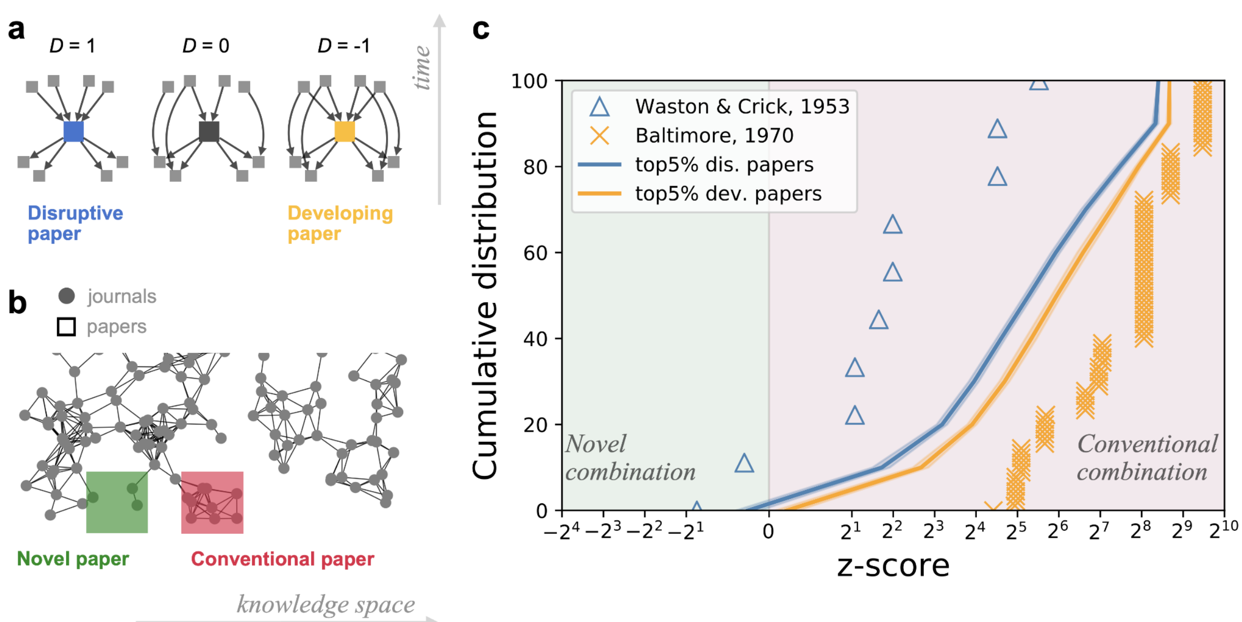

Figure 4. Novel papers disrupt, conventional papers develop. (a) Simplified illustration of disruption. Three citation networks comprising focal papers (colored squares), references and subsequent work (gray squares). Subsequent work may cite the focal work (*i*), both the focal work and its references (*j*) or just its references (*k*). Disruption, *D*, of the focal paper is defined by the difference between the proportion of type *i* and *j* papers $pi - pj$, which equals the difference between the observed number of these papers $ni - nj$ divided by the number of all subsequent works $ni + nj + nk$. A paper may be disrupting ($D = 1$), neutral ($D = 0$) or developing ($D = -1$). Figure recreated from (Wu, Wang, and Evans 2019). (b) Simplified illustration of novelty. A paper may cite journals of weak (green, $z < 0$) or strong ties (red, $z > 0$). (c) Cumulative distributions of z-scores for two

exemplary papers: the DNA paper by Waston and Crick (D = 0.96, top 1% disruptive) and the RNA paper by Baltimore (D = -0.47, top 1% developing). For all 87,475 papers published in 1970, we selected the most disruptive (top 5%) and developing (top 5%) papers, then calculated their average cumulative distribution of conventionality (displayed in blue and orange, respectively). Median z-scores for disruptive papers are significantly different from those of developing papers (the Kolmogorov–Smirnov statistic $D = 0.14$, $p < 0.001$). The same conclusion holds for the 10 percentile z-score ($D = 0.09$, $p < 0.001$). Note that Panel c is plotted using the "symlog" (which means symmetrical log) function from the "matplotlib" library in Python. It displays positive and negative values in the same axis, by setting a small range around zero on the x-axis as linear instead of logarithmic, which explains why zero appears on the x-axis in log scale.

Fig.4 presents the association between novelty and disruption with two representative cases after illustrating the calculation of D- and z-scores. In Fig. 4c, each paper is defined as a random sample of journal co-citations, whose z-score distribution characterizes paper novelty. A high, positive z-score (the distribution shifting to the right end on the x-axis) is a signature of typical combinations of journals on established topics within a field, whereas a low, negative z-score (the distribution shifting to the left end on the x-axis) implies unusual combinations of journals that span fields. The average z-score distribution of the most disruptive (top 5% D-score) versus developing (bottom 5% D-score) papers significantly deviate from one another as evidenced by Kolmogorov-Smirnov tests (see the caption of Fig.4). The former shifts to the left and the latter shifts to the right, indicating the alignment between novelty and disruption; conventionality and development.

The observed association between novelty and disruption raises the question of how it enriches the picture of scientific advance in which novelty, citation impact, and disruption interplay. Scientific advance is constrained by the essential tension between "relevance vs. originality" (Whitley 2000). New ideas need to be introduced to the main body of scientific knowledge based on their relevance. This permits two types of strategies for individual scientists to effectively translate creativity into progress. One can prioritize relevance by selecting an established topic and maximizing originality to bridge knowledge into established fields. This strategy, characterized by the "flat" cumulative distribution (Figure S1) in z-score space harvests mature attention to make a "hit" paper (Brian Uzzi et al. 2013). Alternatively, one can prioritize originality by selecting an underdeveloped topic lacking consensus. In this case, one needs to present discovered relevance between cited ideas and references as critical to the new insight. In the space of z-scores, this strategy corresponds to shifting the cumulative distribution left, which results in a higher likelihood of being disruptive.

To obtain a more intuitive understanding of these complex dynamics characterizing creative destruction in science (McMahan and McFarland 2021), we highlight two papers, one on the double helix structure of DNA by Waton and Crick (Watson and Crick 1953) called the "DNA" paper hereafter, and another on "RNA-dependent DNA Polymerase" by David Baltimore called the "RNA" paper hereafter. The two papers are similar in many ways: both are highly-cited, field-definitive work by distinguished biologists later awarded the Nobel Prize in Physiology or Medicine. However, the z-score distributions reveal their distinct approaches to integrate prior knowledge (Fig. 4c). Baltimore's paper reviewed papers published in conventional biology venues ($z_{median} = 266.3$), including *Virology* versus *itself* ($z = 710.6$), *Virology* versus *Proceedings of the National Academy of Sciences of the United States of America* ($z = 266.3$), and hypothesized that genetic information could transfer bidirectionally between DNA and RNA. At the time of this paper, information transfer from DNA to RNA was well studied, and Baltimore was not the only scholar proposing to test the reverse influence from RNA to DNA. Actually, Baltimore's paper was published back-to-back with Howard Temin's paper (Mizutani, Boettiger, and Temin 1970) on the same topic in the same issue of *Nature*. The bidirectional influence between DNA and RNA represents the "adjacent possible" described by Kauffman (Kauffman 1996), which suggests new ideas or discoveries that extend from prior science, a single step from present understanding as "low hanging fruit", easily reached, which, unsurprisingly, triggers intense competition. The back-to-back publications by Baltimore and Temin were like the race to the South Pole between Britain's explorer Robert Scott and Norway's Roald Amundsen. Unlike the explorers' race, which ended in victory for Amundsen and tragedy for Scott, the discovery of DNA-RNA's mutual influence became a shared and widely celebrated success treated as a confirmation of the underlying claim. In 1975, only five years after the paper's publication, Baltimore and Temin shared the Nobel Prize in Physiology or Medicine. This timely appreciation itself speaks for the adjacent, developing nature of that discovery, evidenced by its low D-score (D = -0.47, bottom 1% disruption, or top 1% development).

In comparison, Watson and Crick's paper cited prior literature published in diverse journals across fields ($z_{median} = 4.8$), including *Journal of Geophysical Research* and *Journal of Chemical Physics* ($z =- 26.7$), *Canadian Journal of Chemistry* and *Quarterly Journal of the Royal Meteorological Society* ($z =- 5.7$), proposing that double-stranded DNA of helical structure is the genetic material. This paper was ahead of time. When published, there was not yet a consensus on the identity of genetic material—proteins seemed a better bet. Moreover, few could foresee its influence into the future; how the double helix shed light on almost every aspect of modern biology and medicine for decades to come, ranging from the migration of human populations and cancer-causing mutations in tumors to the diagnosis and treatment of rare congenital diseases. Watson and Crick received delayed recognition of the Nobel Prize in Physiology or Medicine in 1962, ten years after the paper was published—an enduring wait twice longer than Baltimore's work despite its greater impact in transforming the future of biology and offering the non-academic world an icon of scientific work—the double helix (Ferry 2019). This delayed acknowledgement footnotes the pioneering, disruptive nature of that discovery, evidenced by its high D-score (D=0.96, top 1% disruptive).

Besides the works with clear tasks and searching for evidence to build or criticize a theory, many studies are filled with ambiguity and the central meanings of topics are modified through the time. New topics are not necessarily the topics that are completely new with different names. New topics are formed when they are defined or redefined. It can have a new name with central meanings or with an old name but with different central meanings. In either way, the formation of a topic needs a community around it and a central paper defining its meanings. For example, the theory: 'The Market for Lemons' that discusses the consequence of information asymmetry on markets between buyers and sellers was published as "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism"in the year of 1970 (Akerlof 1970). Empirically, this work looks at the papers that were published in the year of 1970, and maps 914 of papers as the central papers of 943 topics among all papers that focus on the same topic. The center of a topic is defined as the most highly cited paper among all papers that focus on the same topic (King, Downey, and Weld 2020). This work treats a paper that focuses on a topic if the text of the topic appears in the title of that paper. Furthermore, the topics are mapped with the 'Fields of Study' file from Microsoft Academic Graph that contains 70 millions of fields of study. This work finds that the papers that are regarded as defining the central meaning of a topic are more disruptive (Fig. 5a) and are more likely to combine atypical combinations of existing knowledge (Fig. 5b).

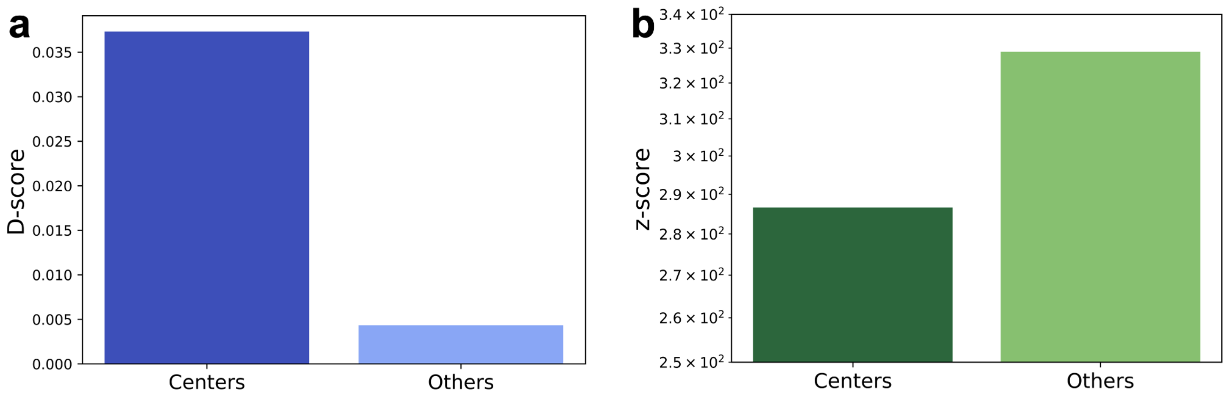

Figure 5. Papers that define or redefine a topic are more likely to be disruptive and combine atypical journal pairs than papers that did not propose new topics. In the year of 1970, 943 topics of study were defined or redefined with the average popularity of topics being 20. Among all 87,475 papers published in 1970, 914 of them are regarded as the central of the 943 topics. (a) The average of D scores of the papers that proposed new topics (0.04) is higher than that of papers that did not (0.004) (t-statistic = -22.49, $p < 0.001$). (b) The average of median z-scores of the papers that proposed new topics (286.45) is higher than that of papers that did not (328.75) (t-statistic = -2.86, $p < 0.05$).

2. Novel papers are more likely to become "sleeping beauties" and accumulate citation impact over the long run

Going back to the time of publication, can one have foreseen the accelerated acknowledgment of Baltimore's contribution and retarded recognition to Watson and Crick? Could we predict them from the conventionally of the former ($z_{median} = 266.3$) and novelty of the latter ($z_{median} = 4.8$), which can be derived at the moment of publication? In these cases, and millions of others published over the following decades, we document that novelty results in delayed impact. Creative explorations that travel more than a step beyond the adjacent possible are inaccessible to the majority of scientists upon publication, but come to make up the pool of possibilities that are verified, appreciated, or reformulated and used to advance science over the long run.

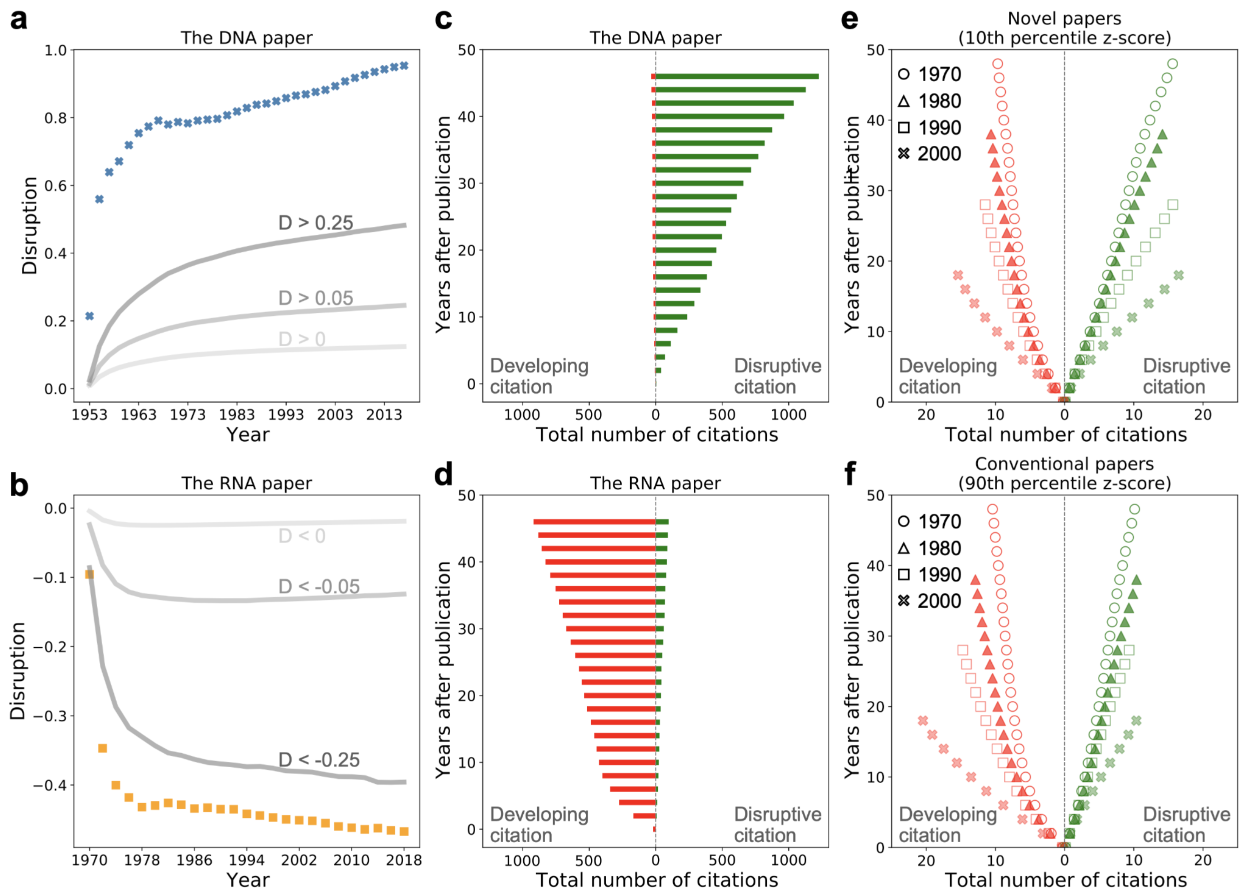

Figure 6. The disruption of novel papers increases over time. The temporal evolution of disruption scores for the DNA paper by Waston and Crick (a) and RNA paper by Baltimore (b). In Panels a-b, each grey line represents a group of papers sharing a similar level of D-score annotated in the figure. In Panels c and d, we display the asymmetry in the total number of disruptive (green) and developing (red) citations to the DNA paper and RNA paper over time, respectively. This analysis is extended to four generations of papers, including the cohort of 1970 (87,475 papers), 1980 (176,826 papers), 1990 (318,914 papers), and 2000 (591,653 papers). For each cohort, we select the most novel (top 10% z-score, Panel e) and conventional papers (bottom 10%, Panel f) and plot the average total number of citations over time. Statistical tests on the asymmetry between the two kinds of citations were reported as follows. The t-test of the difference between novel and conventional papers in disruptive-citation fraction at the last year of analysis (2018) is significant for all cohorts, including 1970 (t-statistic = 23.85, $p < 0.001$), 1980 (t-statistic = 33.04, $p < 0.001$), 1990 (t-statistic = 52.78, $p < 0.001$), 2000 (t-statistic = 88.76, $p < 0.001$).

We examined the temporal evolution of disruption and citations of these two, archetypal papers (Fig. 6a-b). We find that the D-score of the DNA paper increased nearly monotonically from 0.2 in 1953 to nearly 0.8 five years after publication, steadily increasing to 0.96 in 2018, whereas the D-score of the Baltimore paper has been negative since publication in 1970, and decayed rapidly to -0.45 within five years of publication, barely changing in the subsequent four decades (D = -0.47 in 2018, bottom 1%). This seems to be a general pattern we confirmed from many other cases: the D-score of developing papers converges quickly within five years, but that of disruptive papers increases after a decade or longer (Fig. S2).

For both the DNA and RNA papers, we unpacked two kinds of citations: disruptive citations from subsequent papers that only cite the focal paper but not its references (green curves in Fig. 6c), and developing citations from subsequent papers that cite both (the red curves in Fig. 6d). We find a "taking off" pattern in the DNA paper—disruptive citations increase steadily following paper publication, deviating from developing citations, which decline exponentially after a short peak and follow the widespread pattern of citation decay (D. Wang, Song, and Barabási 2013). Disruptive citations contribute to long-term impact more than developing citations. In comparison, citation impact of the RNA paper is increasingly dominated by developing citations, reflecting the stabilizing consensus on its developing contribution within biology .

Long-term impact for novel papers is confirmed as a general pattern when we scale the data (Fig. 6e-f). We select the most novel (top 10%) and conventional (bottom 10%) papers by z-score including the cohort of papers published in 1970 (87,475 papers), 1980 (176,826 papers), 1990 (318,914 papers), and 2000 (591,653 papers then compared the difference between disruptive (green data points in Figure 6e-f) and developing (red data points in Figure 6e-f) citations. Novel papers that integrate surprising combinations of literature to create new ideas accumulate long-term impact by attracting both disruptive and developing citations, with the relative fraction of the former over the latter amplifying over time (Fig. 6e). This pattern reverses in the citation dynamics of conventional papers, wherein the relative fraction of developing citations increases faster than that of disruptive citations (Fig. 6f).

To verify the long-term impact of a novel paper, we calculate the Sleeping Beauty Index (SBI), a citation-based measure designed to capture the convexity of a citation curve over time (Ke et al. 2015). A paper with a high SBI will receive few citations upon publication, followed by a later burst tracing a convex curve. By contrast, a paper with a low SBI will receive many citations following publication and fewer later tracing a concave cumulative distribution. Novelty and SBI are positively correlated (Pearson correlation coefficient equals 0.08 on the log-log scale, p-value < 0.001). We also calculated SBI over time to examine the chance that larger bursts occur after a long wait, i.e., whether the long-term citations compensate for lengthy waiting times and drive up SBI over time. We find that curves of conventional papers flatten within a decade, implying no delayed burst of citation attention. In contrast, the citation curves of novel papers continuously increase after a decade, revealing long-term citation pay-off after lengthy waiting times (Fig. S3).

The correlation between novelty and disruption holds for a majority of fields, but the effect is more significant in "artificial" than "natural" sciences (Simon 2019) In computer science and engineering (the higher average D-score as presented in Fig. 3 in (Wu, Wang, and Evans 2019)), novelty is more likely rewarded in disruption, evidenced by the larger difference between novel versus conventional groups in the fraction of disruptive citations (Table S1). On the contrary, for fields such as biology, chemistry, and physics, which are already hard to disrupt (low average D-score), novelty is more weakly related to disruption.

3. Reformulating novelty as distance in knowledge space to map the moving frontier of science

Z-score based novelty can be reformulated for computational efficiency and dynamism as the distance between journals in embedding spaces built from co-cited journals (see Methods for details). The temporal evolution of these knowledge embedding spaces reveal how yesterday's novelty forms today's scientific conventions, which tomorrow's breakthroughs disrupt.

As described above, z-score can be understood as a measure of the conceptual "tie strength" (Granovetter 1973) between sources, which differentiates "strong ties" within scientific disciplinary communities from "weak ties" between them. Newman and colleagues (Newman 2006) developed "Q-modularity" based on an alternative, analytical version of the z-score to identify network communities, where negative values ($z < 0$) indicate gaps between communities. In other words, network communities are defined as groups of nodes connected by paths of strong ties. Shwed and Bearman measure used Q-modularity on citation networks to quantify the level of scientific consensus on topics such as whether smoking is cancerous (2010). From this perspective, if we construct a latent space of knowledge by reformulating z-score as distance, we can model the conventions and novelties in science as claims within "dense" and "sparse" regions of knowledge space, respectively. This map directs human creativity towards scientific frontiers at the "sparse", novel regions.

Recent advances in natural language processing (NLP) for semantic analysis, such as the word2vec manifold learning model (Mikolov et al. 2013), provides us the tools needed to reconstruct knowledge spaces. With them, we can extend the z-score as a measure of continuous distance across embedding space. Levy and Goldberg analytically prove that PMI, a revised z-score, equals the distance between vectorized items embedded in latent spaces as calculated by inner product (Levy and Goldberg 2014). In this way, knowledge embedding spaces learn scientific conventions, which can be used to assess and direct exploration of the scientific frontier (Tshitoyan et al. 2019).

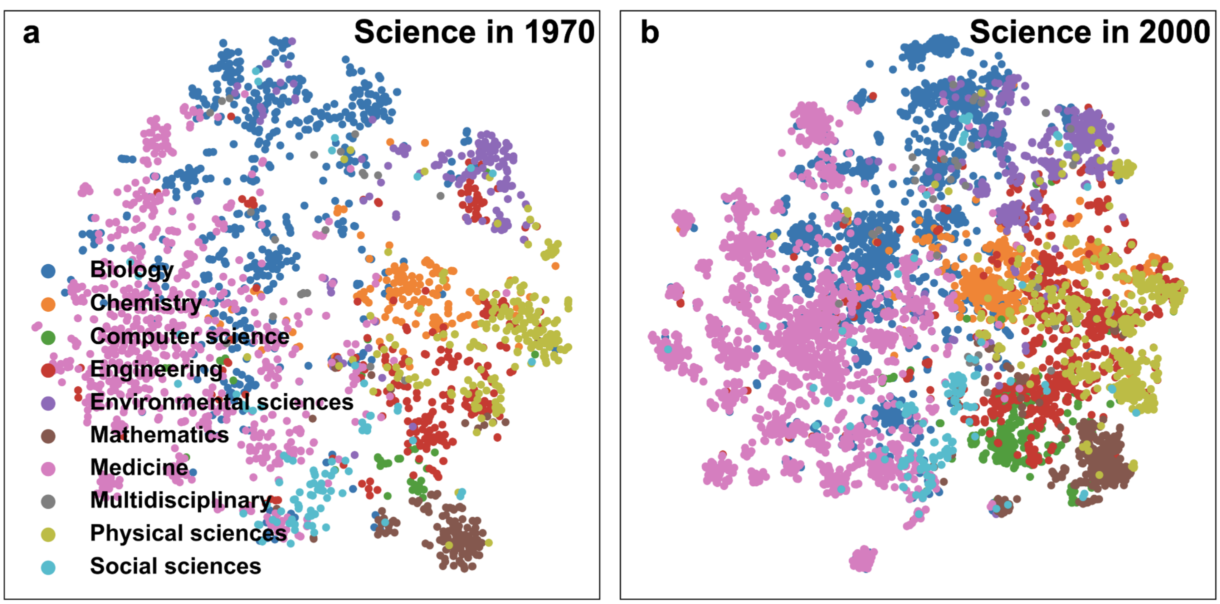

Figure 7. Knowledge spaces in 1970 and 2000 obtained through journal embeddings. We constructed two journal embeddings using the 1970 (2,429 journals) and 2000 (8,009 journals) cohort of papers (see Methods for details of the embeddings). Each dot is a journal colored by field. We trained the embeddings using the word2vec Skip-gram algorithm. We used the Gensim package in Python with parameters as follows: embedding dimension =

50, negative sampling size = 5, and window size = 10. We then project the 50-D journal vectors into a 2-D space using the t-SNE algorithm.

To test the association between z-score and embedding space distance, we constructed two embedding spaces using the journal co-citation networks in 1970 (2,429 journals) and 2000 (8,009 journals), respectively. We also visualize these embedding spaces by projecting it onto a 2-dimension space with the t-SNE algorithm. We find that the distance between journals in the embedding space reflects their content relevance, as journals from the same field tend to cluster together (Fig. 7 and Fig. S4). This observation is confirmed as we zoom-in with a focus on the regions covered by Mathematics and Computer Science journals. While close to each other, the average distance between journals across the two fields is larger than within a field. The z-score for journal pairs correlates strongly with the inner product between journal embeddings (Pearson correlation coefficient = 0.74, p-value < 0.001), confirming the validity of z-score reformutating (Fig. S5).

The landscape of novelty is indeed changing, as revealed by the comparison between knowledge spaces in 1970 and 2000 (Fig. 7). One of the most strikingly visible trends is the emergence of dense areas of journals within each field, suggesting the formation of subfields supported by consensus on relevant topics. In parallel to clustering within fields, fields also mix with one another, showing the increasing importance of interdisciplinary scientific collaboration (Leydesdorff and Ivanova 2021; Leahey, Beckman, and Stanko 2017). The change of relative distance between fields also reveals rich trends in shifting novelty. In 1970s, a study drawing up social science and computer science was highly novel due to the distance between these two fields, but is far less so in 2000, when these two field are closer to each other after waves of movements to link them, including “social informatics” of the 1980s (Kling 1999) and “computational social science” of the past decade (D. Lazer et al. 2009; D. M. J. Lazer et al. 2020).

Section 5: Discussions

We find that D score prefers different types of new ideas compared with the number of citations. More novelty is accepted under the measurement of disruption to make a new direction in science. Although the number of citations suggest there exists bias towards being novel, our results suggest D score is able to signify the formation of a new community and this measurement can get rid of the tendency bias to some degree. This leads us to ask a further question: can innovation be predicted? Recent study on machine learning suggests that the high impact work could be predicted at an early stage. The new characteristics of materials could be predicted by the distance on the embedding space with already known materials. AlphaFold solves the puzzle of the structure of protein efficiently. Machines show their power on using the represented knowledge in a powerful way. However, at the same time, whether innovation in science can be predicted is still in debates. Our work shows that when innovation is defined as how it builds up or eclipse the previous works, the atypical combinations contribute to the disruption in the future. Besides, experts believe that the excellence of work can only be evaluated by experts in the domains (Abbott 2014). In this way, our results complement the innovation picture by suggesting that science proceeds through a more serendipitous way. For science, we can imagine there exist a lot of eureka moments and 'you know it when you see it'.

Further research starting from this work can explore better representation of the unstructured knowledge. At first, future research can search for a more accurate representation of conventionality and novelty. The present landscape-based knowledge embedding is not an efficient representation of context-sensitive detection of novelty. Second, all new ideas have an origin and it matters how it is combined. Some knowledge like tacit knowledge should be taken into account during the making of new ideas since tacit knowledge matters when dealing with lack of consensus and unstructured knowledge.

References

Abbott, Andrew. 2014. "The Excellence of IT: Conceptions of Quality in Academic Disciplines." In *The Institution of Science and the Science of Institutions*, 147–65. Dordrecht: Springer Netherlands.

Ahuja, Gautam, Curba Morris Lampert, and Vivek Tandon. 2014. "Paradigm-Changing vs. Paradigm-Deepening Innovation: How Firm Scope Influences Firm Technological Response to Shocks." *Organization Science* 25 (3): 653–69.

Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *The Quarterly Journal of Economics*. https://doi.org/10.2307/1879431.

Baltimore, David. 1970. "Viral RNA-Dependent DNA Polymerase: RNA-Dependent DNA Polymerase in Virions of RNA Tumour Viruses." *Nature* 226 (5252): 1209–11.

Bourdieu, Pierre. 1975. "The Specificity of the Scientific Field and the Social Conditions of the Progress of Reason." *Social Sciences Information. Information Sur Les Sciences Sociales* 14 (6): 19–47.

Burt, Ronald S. 2007. *Brokerage and Closure: An Introduction to Social Capital*. OUP Oxford.

Chen, Jiyao, Diana Shao, and Shaokun Fan. 2021. "Destabilization and Consolidation: Conceptualizing, Measuring, and Validating the Dual Characteristics of Technology." *Research Policy* 50 (1): 104115.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” *arXiv [cs.CL]*. arXiv. http://arxiv.org/abs/1810.04805.

Dosi, Giovanni. 1982. “Technological Paradigms and Technological Trajectories: A Suggested Interpretation of the Determinants and Directions of Technical Change.” *Research Policy* 11 (3): 147–62.

Ferry, Georgina. 2019. “The Structure of DNA.” *Nature* 575 (7781): 35–36.

Foster, Jacob G., Andrey Rzhetsky, and James A. Evans. 2015. “Tradition and Innovation in Scientists’ Research Strategies.” *American Sociological Review* 80 (5): 875–908.

Funk, Russell J., and Jason Owen-Smith. 2016. “A Dynamic Network Measure of Technological Change.” *Management Science* 63 (3): 791–817.

Garud, Raghu, Arun Kumaraswamy, and Peter Karnøe. 2010. “Path Dependence or Path Creation?” *The Journal of Management Studies* 47 (4): 760–74.

Granovetter, Mark S. 1973. “The Strength of Weak Ties.” *The American Journal of Sociology* 78 (6): 1360–80.

Grover, Aditya, and Jure Leskovec. 2016. “node2vec: Scalable Feature Learning for Networks.” *KDD: Proceedings / International Conference on Knowledge Discovery & Data Mining. International Conference on Knowledge Discovery & Data Mining* 2016 (August): 855–64.

Hicks, Diana, Paul Wouters, Ludo Waltman, Sarah de Rijcke, and Ismael Rafols. 2015. “Bibliometrics: The Leiden Manifesto for Research Metrics.” *Nature* 520 (7548): 429–31.

Hofstra, Bas, Vivek V. Kulkarni, Sebastian Munoz-Najar Galvez, Bryan He, Dan Jurafsky, and Daniel A. McFarland. 2020. “The Diversity–Innovation Paradox in Science.” *Proceedings of the National Academy of Sciences of the United States of America* 117 (17): 9284–91.

Karim, Samina, and Will Mitchell. 2017. “Path-Dependent and Path-Breaking Change: Reconfiguring Business Resources Following Acquisitions in the US Medical Sector, 1978-1995.” *The SMS Blackwell Handbook of Organizational Capabilities*. https://doi.org/10.1002/9781405164054.ch13.

Kauffman, Stuart. 1996. *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*. Oxford University Press.

Ke, Qing, Emilio Ferrara, Filippo Radicchi, and Alessandro Flammini. 2015. “Defining and Identifying Sleeping Beauties in Science.” *Proceedings of the National Academy of Sciences of the United States of America* 112 (24): 7426–31.

King, Daniel, Doug Downey, and Daniel S. Weld. 2020. “High-Precision Extraction of Emerging Concepts from Scientific Literature.” In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM. https://doi.org/10.1145/3397271.3401235.

Kling, Rob. 1999. “What Is Social Informatics and Why Does It Matter?” *D-Lib Magazine*. https://doi.org/10.1045/january99-kling.

Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers Through Society*. Harvard University Press.

Lazer, David M. J., Alex Pentland, Duncan J. Watts, Sinan Aral, Susan Athey, Noshir Contractor, Deen Freelon, et al. 2020. “Computational Social Science: Obstacles and Opportunities.” *Science* 369 (6507): 1060–62.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-Laszlo Barabasi, Devon Brewer, Nicholas Christakis, et al. 2009. “Social Science. Computational Social Science.” *Science* 323 (5915): 721–23.

Leahey, Erin, Christine M. Beckman, and Taryn L. Stanko. 2017. “Prominent but Less Productive: The Impact of Interdisciplinarity on Scientists’ Research.” *Administrative Science Quarterly* 62 (1): 105–39.

Levy, Omer, and Yoav Goldberg. 2014. “Neural Word Embedding as Implicit Matrix Factorization.” In *Advances in Neural Information Processing Systems 27*, edited by Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, 2177–85. Curran Associates, Inc.

Leydesdorff, Loet, and Inga Ivanova. 2021. “The Measurement of ‘interdisciplinarity’ and ‘synergy’ in Scientific and Extra‐scientific Collaborations.” *Journal of the Association for Information Science and Technology* 72 (4): 387–402.

March, James G. 1991. “Exploration and Exploitation in Organizational Learning.” *Organization Science* 2 (1): 71–87.

McMahan, Peter, and Daniel A. McFarland. 2021. “Creative Destruction: The Structural Consequences of Scientific Curation.” *American Sociological Review* 86 (2): 341–76.

Merton, Robert K. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press.

Miao, Lili, Dakota Murray, Woo-Sung Jung, Vincent Larivière, Cassidy R. Sugimoto, and Yong-Yeol Ahn. 2021. “The Latent Structure of National Scientific Development.” *arXiv [cs.SI]*. arXiv.

http://arxiv.org/abs/2104.10812.
Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." *arXiv [cs.CL]*. arXiv. http://arxiv.org/abs/1301.3781.
Mizutani, Satoshi, David Boettiger, and Howard M. Temin. 1970. "A DNA-Dependent DNA Polymerase and a DNA Endonuclease in Virions of Rous Sarcoma Virus." *Nature* 228 (5270): 424–27.
Newman, M. E. J. 2006. "Modularity and Community Structure in Networks." *Proceedings of the National Academy of Sciences of the United States of America* 103 (23): 8577–82.
Peng, Hao, Qing Ke, Ceren Budak, Daniel M. Romero, and Yong-Yeol Ahn. 2020. "Neural Embeddings of Scholarly Periodicals Reveal Complex Disciplinary Organizations." *arXiv [cs.DL]*. arXiv. http://arxiv.org/abs/2001.08199.
Perozzi, Bryan, Rami Al-Rfou, and Steven Skiena. 2014. "Deepwalk: Online Learning of Social Representations." In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 701–10.
Polanyi, Michael. 1962. "The Republic of Science." *Minerva* 1 (1): 54–73.
Shwed, Uri, and Peter S. Bearman. 2010. "The Temporal Structure of Scientific Consensus Formation." *American Sociological Review* 75 (6): 817–40.
Simon, Herbert A. 2019. *The Sciences of the Artificial*. MIT Press.
Tshitoyan, Vahe, John Dagdelen, Leigh Weston, Alexander Dunn, Ziqin Rong, Olga Kononova, Kristin A. Persson, Gerbrand Ceder, and Anubhav Jain. 2019. "Unsupervised Word Embeddings Capture Latent Knowledge from Materials Science Literature." *Nature* 571 (7763): 95–98.
Tushman, Michael L., and Philip Anderson. 1986. "Technological Discontinuities and Organizational Environments." *Administrative Science Quarterly* 31 (3): 439–65.
Uzzi, B., S. Mukherjee, M. Stringer, and B. Jones. 2013. "Atypical Combinations and Scientific Impact." *Science*. https://doi.org/10.1126/science.1240474.
Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Ben Jones. 2013. "Atypical Combinations and Scientific Impact." *Science* 342 (6157): 468–72.
Wang, Dashun, Chaoming Song, and Albert-László Barabási. 2013. "Quantifying Long-Term Scientific Impact." *Science* 342 (6154): 127–32.
Wang, Jian, Reinhilde Veugelers, and Paula Stephan. 2016. "Bias against Novelty in Science: A Cautionary Tale for Users of Bibliometric Indicators." https://doi.org/10.3386/w22180.
Watson, J. D., and F. H. Crick. 1953. "Molecular Structure of Nucleic Acids; a Structure for Deoxyribose Nucleic Acid." *Nature* 171 (4356): 737–38.
Whitley, Richard. 2000. *The Intellectual and Social Organization of the Sciences*. Oxford University Press.
Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi. 2007. "The Increasing Dominance of Teams in Production of Knowledge." *Science* 316 (5827): 1036–39.
Wu, Lingfei, Dashun Wang, and James A. Evans. 2019. "Large Teams Develop and Small Teams Disrupt Science and Technology." *Nature* 566 (7744): 378–82.
Xian, Yongqin, Zeynep Akata, Gaurav Sharma, Quynh Nguyen, Matthias Hein, and Bernt Schiele. 2016. "Latent Embeddings for Zero-Shot Classification." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 69–77.
Xie, Huang, and Tuomas Virtanen. 2019. "Zero-Shot Audio Classification Based On Class Label Embeddings." In *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 264–67.

Supplement Materials

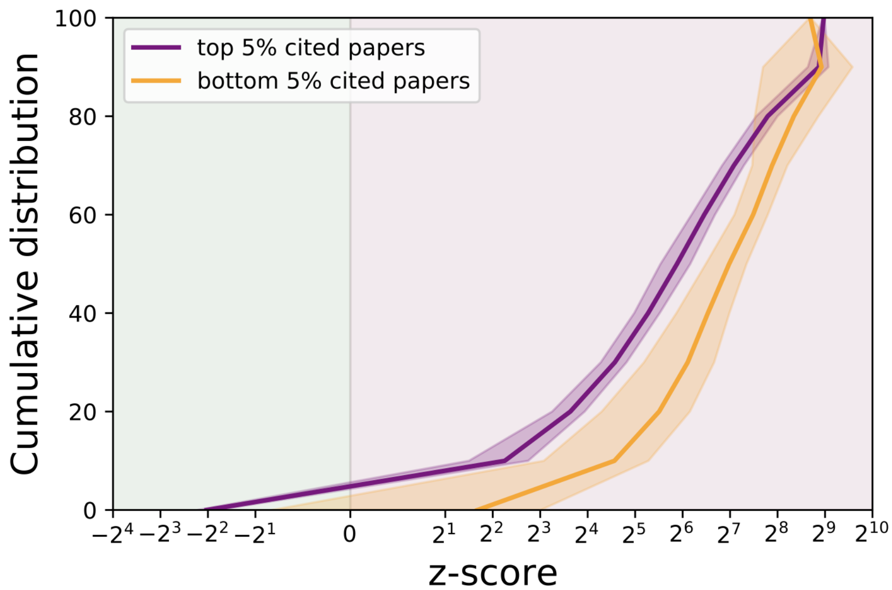

Figure S1. High-impact papers mix novel and conventional combinations of literature in the references. The cumulative distributions of the z-score of high-impact (purple, top 5% citations) and low-impact (yellow, bottom 5% citations) papers selected from all the 87,475 papers published in 1970. The median z-score of high-impact papers is significantly different from that of low-impact papers (the Kolmogorov–Smirnov statistic $D = 0.15$, $p < 0.001$). Note that Figure S1 is plotted using the "symlog" (which means symmetrical log) function from the "matplotlib" library of Python. It displays positive and negative values in the same axis, by setting a small range around zero on the x-axis as linear instead of logarithmic, which explains why zero appears on the x-axis of logarithmic scales.

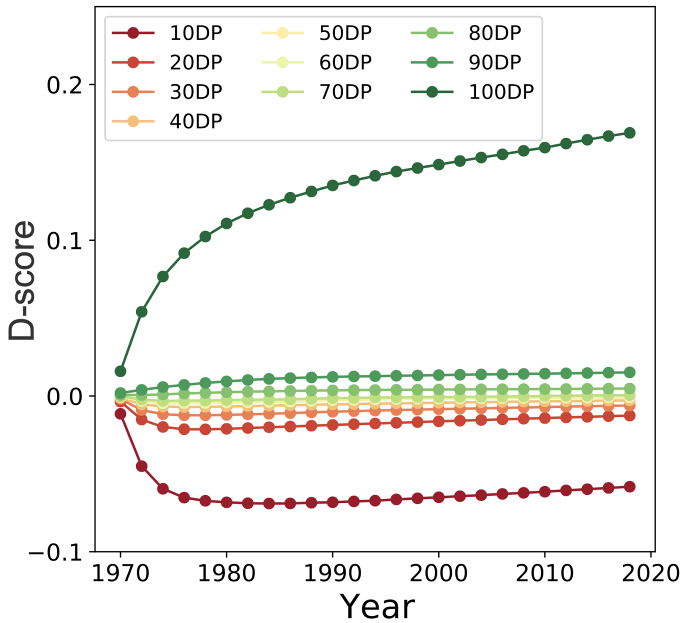

Figure S2. D-score converges fast for developing papers and slow for disruptive papers. From all the 87,475 papers published in 1970, we break them into ten groups by disruption percentiles (DP) and plot the average disruption score of papers within the group against years. The curves are colored by DP.

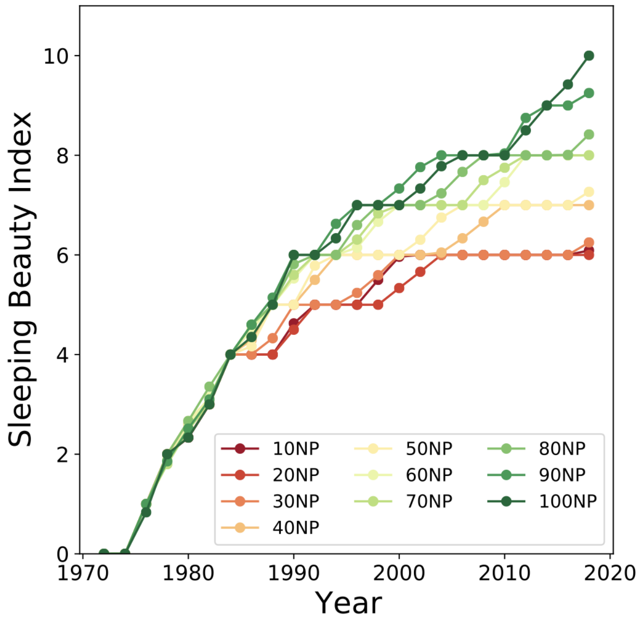

Figure S3. Novel papers are more likely to become "sleeping beauties" and the long-term citations pay off the waiting time. From all the 87,475 papers published in 1970, we break them into ten groups by novelty percentiles (NP) using z-score, and plot the median Sleeping Beauty Index (SBI) of papers within the group against years. In the original version, each paper has only one SBI calculated over the lifecycle of citations (Ke et al. 2015). Here we calculate and track the temporal evolution of SBI for each two-year time window. The average curves of low-novelty papers are flatten within a decade, showing that no delayed burst of citations to the papers. The average curves of high-novelty papers continuously increase after a decade, showing that new, larger bursts still happen after a long wait, i.e., the long-term citations pay off the lengthy waiting time.

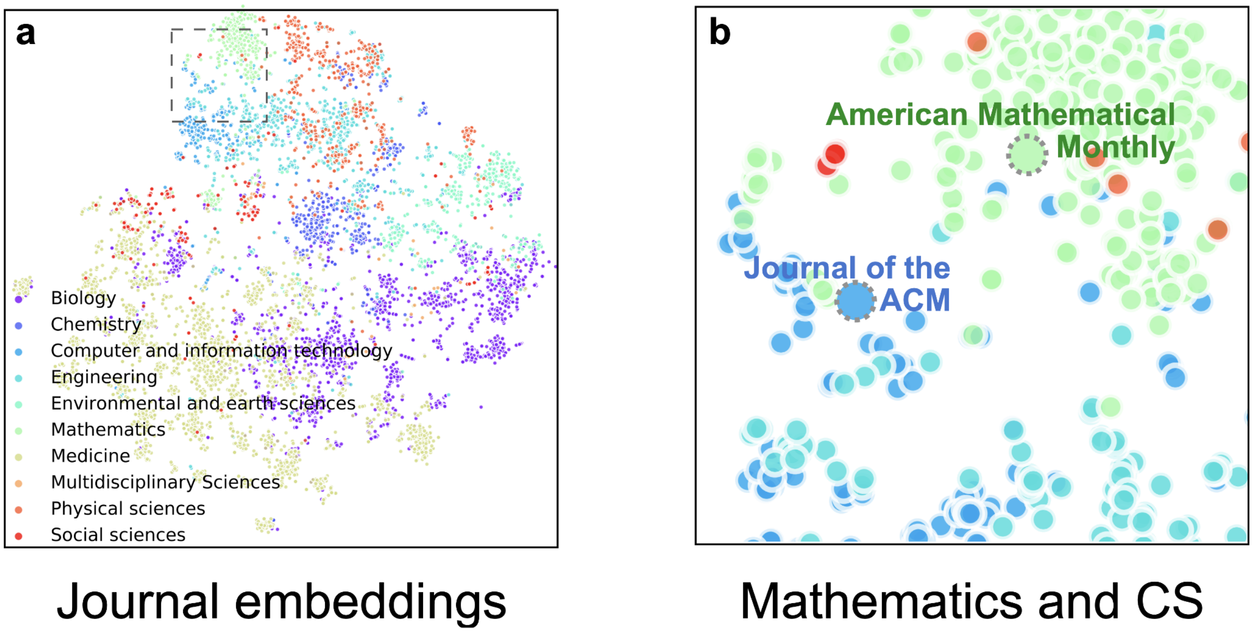

Figure S4. Inner products in a journal embedding space are a proxy for z-score novelty. (a) A journal embedding space constructed using the 2000 publications dataset (see methods for details), with journals (dots) colored by fields (using a different color scheme from Fig. 6). (b) A zoom-in view of (a) highlighting Mathematics (green dots) and Computer Science journals (blue dots).

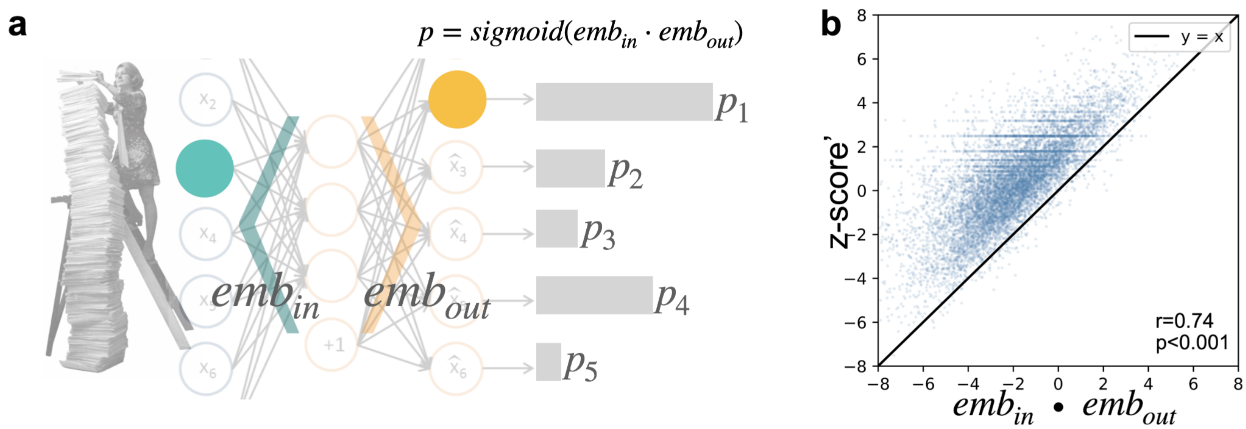

Figure S5. Inner products in a journal embedding space are a proxy for z-score novelty. (a)Illustration of prediction made by skip-gram with negative sampling of word2vec after the training process. Given a center word(green), the trained model returns the probability *p* of its context over the whole vocabulary. *p* is modeled as the inner product between in-embedding of the center word and out-embedding of the context word over the sigmoid activation function. (b) Relationship between symmetric inner dot product and the revised z-score (see Eq. 3). The Pearson correlation coefficient is 0.74 and the p-value is much smaller than 0.001.

Table S1. The asymmetry between disruptive and developing citations conditioned by paper novelty holds across disciplines. For the 87,475 papers published in 1970, 45,677 (52%) among them are matched with fields of study provided by Microsoft Academic Graph. We select 11 fields of 100 or more papers. For each field we select the most novel (bottom 10 z-score) and conventional papers (top 10 z-score). For each paper in these two groups, we calculate the fraction of disruptive citations in the last year of analysis (2018). We then apply a statistical test (t-test) on the difference between the means of the average disruptive-citation fraction between the two groups.

Field	N of papers	N of novel papers (bottom 10% z-score)	N of conventional papers (top 10% z-score)	Average of disruptive-citation fraction at the paper level in the novel group	Average of disruptive-citation fraction at the paper level in the conventional group	The difference between novel and conventional papers in the fraction of disruptive citations	T-test of the difference
Biology	12587	1259	1240	0.508	0.395	0.113	10.09***
Chemistry	12052	1205	1204	0.519	0.512	0.007	0.62
Medicine	6253	624	627	0.554	0.454	0.100	5.99***
Physics	5298	529	313	0.492	0.398	0.094	4.39***
Psychology	3139	314	331	0.496	0.336	0.160	6.51***
Mathematics	2491	247	249	0.540	0.430	0.110	4.00***
Materials Science	1435	144	142	0.549	0.469	0.080	2.29*
Geology	986	99	97	0.547	0.429	0.118	2.87**
Computer Science	398	40	34	0.614	0.399	0.215	3.07**
Engineering	273	27	31	0.654	0.356	0.298	4.44***
Economics	270	27	26	0.722	0.373	0.349	3.95***
Sociology	267	27	26	0.750	0.541	0.209	2.66*

* p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.