Journal of Informetrics 18 (2024) 101594



Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi





On the temporal diversity of knowledge in science

Alex J. Yang *

School of Information Management, Nanjing University, Nanjing, 210023, PR China

ARTICLE INFO

Keywords:
Temporal diversity
Science of science
Citation impact
Scientific innovation

ABSTRACT

Understanding the diversity of scientific knowledge is pivotal for elucidating trends in science and innovation. While interdisciplinarity and team diversity have been extensively studied, the temporal diversity of knowledge remains underexplored. This paper introduces a novel framework for assessing temporal diversity in scholarly research. Analyzing 31 million articles from the past seven decades, I revealed an increasing trend in temporal diversity, reflecting the cumulative nature of scientific knowledge. Additionally, I found that temporal diversity is negatively associated with citation impact but positively associated with disruption. These patterns are robust and consistent across different contexts. Moreover, the findings suggest that higher temporal diversity leading to greater disruption may be primarily due to the use of older references. However, the disadvantages of temporal diversity in terms of citation impact cannot be entirely explained by other factors. Collectively, this study elucidates the dynamics of temporal diversity and its implications for innovation, providing new frameworks in the science of science and evidence on how innovation is driven by the temporal diversity of knowledge.

1. Introduction

The pursuit of scientific knowledge is dynamic and multifaceted, spanning diverse disciplines, perspectives, and temporal dimensions (Gates et al., 2019; Mukherjee et al., 2017; Stirling, 2007). Central to this pursuit is the concept of knowledge diversity, which includes not only the integration of different disciplines (interdisciplinarity) but also the temporal distribution of knowledge across epochs (temporal diversity). While interdisciplinarity and team diversity have received considerable attention in academic discourse (Kwon et al., 2020; Leydesdorff et al., 2019; Li et al., 2023; Song et al., 2023; Sun et al., 2021; Vantard et al., 2023; Zhang et al., 2024), the exploration of temporal diversity in scholarly research remains relatively nascent. Temporal diversity refers to the diversity of scientific knowledge across different temporal epochs, reflecting the evolution and dissemination of ideas over time (Mukherjee et al., 2017; Zeng et al., 2017). Despite its crucial role in shaping scholarly discourse, temporal diversity of knowledge has received limited attention in empirical research.

Understanding the intricacies of temporal diversity of knowledge is crucial for unraveling the trends shaping scientific innovation (Fortunato et al., 2018; Liu et al., 2023; Zeng et al., 2017). This concept is linked to Newton's famous assertion, "If I have seen further than others, it is by standing upon the shoulders of giants" (Furman & Stern, 2011), and also relates to Robert Merton's theory of the "births of time" (Merton, 1961). Such perspectives compel us to delve deeper into the role of temporal diversity of knowledge in shaping innovation in science. Scholars have long recognized the importance of interdisciplinarity in driving innovation and addressing complex challenges that transcend traditional boundaries (Zhang et al., 2024). However, the temporal dimension of

* Correspondence at: School of Information Management, Nanjing University, Nanjing, 210023, PR China.
E-mail address: alexjieyang@smail.nju.edu.cn.

https://doi.org/10.1016/j.joi.2024.101594
Received 31 March 2024; Received in revised form 9 September 2024; Accepted 19 September 2024
1751-1577/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

scientific knowledge remains underexplored. This paper seeks to bridge this gap by introducing a novel framework for assessing temporal diversity in scholarly research and examining its dynamics and associations with scientific impact and innovation over the past seven decades. Building upon these considerations, this study addresses the following research questions:

(RQ1) What are the dynamic trends of temporal diversity in scientific papers over the past 70 years?
(RQ2) What is the relationship between the temporal diversity and scientific impact? Additionally, what is the relationship between temporal diversity and disruptive potential?
(RQ3) To what extent is the observed relationship between temporal diversity, scientific impact, and disruption attributable to reference age? Are there other contributing factors, such as interdisciplinarity, atypicality, or delayed citation, that influence these relationships?

This study contributes to the literature by introducing an innovative framework to quantify the temporal diversity of knowledge and conducting a comprehensive analysis of its dynamic trends, as well as its associations with scientific impact and disruptive potential (Funk & Owen-Smith, 2017; Park et al., 2023). Utilizing an extensive dataset comprising over 31 million journal articles spanning from 1950 to 2016 across all fields, this paper uncovers an increasing trend in temporal diversity, reflecting the cumulative nature of scientific knowledge. The findings reveal that temporal diversity is negatively associated with citation impact while positively correlated with disruptive potential. These patterns are robust and consistent across various contexts. Ultimately, this research enhances the understanding of the complex interrelationships within scientific temporal diversity, informing scholarly discourse, research policy, and practice, and advancing knowledge production and innovation in scientific research.

The structure of this paper is as follows. Section 2 presents a comprehensive review of the existing literature. Section 3 introduces the methods, framework, and details the data utilized in the empirical analysis. Section 4 presents the analysis of the obtained results. Finally, Section 5 delves into the implications of the findings and suggesting avenues for future research.

2. Literature review

Theories of knowledge development emphasize the significance of past information in the generation of new ideas (Lin et al., 2022; Uzzi et al., 2013). However, these theories diverge on where to search for the most fruitful information. One school of thought posits that older work, having withstood the test of time, is most likely to provide the foundational elements for new research. This notion is encapsulated in Isaac Newton's famous assertion, "If I have seen further than others, it is by standing upon the shoulders of giants" (Furman & Stern, 2011). Conversely, the "births of time" theory (Merton, 1961) argues that recent information is more likely to drive high-impact and high-fitness ideas (Wang et al., 2013).

Prior studies have illuminated the dynamics of knowledge formation in science, revealing a consistent increase in both the volume of literature and the average number of references over time, thereby contributing to the accumulation of knowledge. Jones (2009); Jones and Weinberg (2011) and Bloom et al. (2020) have argued that this accumulation of knowledge concomitantly increases the burden of innovation. In alignment with Merton's theory (Merton, 1961), many knowledge and literature retrieval systems (e.g., Web of Science and Google Scholar) prioritize recent information. Cognitively, individuals tend to recall the most recent information first, and search engines typically rank results based on either recency or popularity, both of which correlate with the age of the knowledge.

An important body of literature relevant to this study examines the relationship between the age of knowledge and future citation impact. Numerous researchers have investigated various aspects of how referencing behavior characteristics, particularly reference age and recency, correlate with citation impact (Herrmannova et al., 2018; Liu et al., 2022; Mammola et al., 2022; Onodera & Yoshikane, 2015). Haslam et al. (2008) demonstrated that articles with newer references tend to receive more citations. Onodera and Yoshikane (2015), in their study involving five natural sciences and one engineering field, identified recency of cited references (measured by the Price Index) as the strongest predictor of citation rate. Ahlgren et al. (2018), through quantile regression analysis, found that scientific publications with references from recent years tend to achieve greater citation impact. Furthermore, Mukherjee et al. (2017) discovered that works citing recent literature generally have higher scientific impact.

Another body of literature addresses the disruptive potential of scientific papers. Disruptive innovation encapsulates pioneering discoveries or methodologies that challenge existing paradigms, marking a departure from incremental progress toward radical shifts in knowledge frameworks (Kuhn, 1962; Merton, 1973). Quantifying disruption in science involves analyzing citation links within the knowledge network, encompassing both the forward and backward impacts of papers, thereby contributing to an extensive knowledge flow network (Yang & Deng, 2024). The CD index, introduced by Funk and Owen-Smith (2017), meticulously assesses the disruptive nature of technologies by examining structural features within the deep citation network. Building upon this index, Wu et al. (2019) have extended its application to scientific literature, providing a quantitative measure for evaluating the disruptive nature of scientific knowledge. There exists contention surrounding claims that innovative activities are slowing down (Bloom et al., 2020; Klüppel & Knott, 2023). Park et al. (2023) demonstrated that both papers and patents exhibit decreased levels of disruption over time. They argue that this trend may be attributable to the increasing burden of knowledge. Notably, the CD index is significantly influenced by the age of references, as older references may lead to a disconnect with future citations. This suggests a potential association between the age of references and the CD index. Additionally, Wu et al. (2019) found that disruption relates to the phenomenon of delayed citation, which can also be influenced by the age of references.

Previous studies have predominantly relied on average reference ages or the Price index (Onodera & Yoshikane, 2015) (e.g., the percentage of references published within five years preceding the publication year of the article) as measures of temporal knowledge. However, these methods are insufficient to capture the full range of temporal distribution in scholarly references. This paper seeks to

address this limitation by introducing a novel framework for assessing temporal diversity in academic research. Furthermore, existing literature on reference ages tends to focus narrowly on specific fields, failing to identify general patterns across disciplines. This study provides a comprehensive analysis, encompassing 31 million papers, to offer broader insights into temporal knowledge distribution.

3. Data

This study leveraged the Microsoft Academic Graph (MAG) dataset, recognized for its extensive coverage and reliability in scholarly research (Wang et al., 2020). It specifically utilized the iteration of the MAG dataset released on December 6th, 2021. Encompassing data from 1800 to 2021, this dataset comprises a vast corpus of over 200 million documents, including journal articles, conference proceedings, preprints, and diverse research publications.

I aimed to conduct a comprehensive analysis of temporal diversity across all domains and years in science, considering the significant shifts in modern science since World War II (Lin et al., 2023; Wu et al., 2019). Additionally, restricting the sample to papers published before 2017 ensures a minimum citation window of five years (Wang, 2013), which is necessary for accurately calculating the primary variables: citation count and the CD index. In order to provide a nuanced understanding of temporal diversity, this analysis thus focused on papers published between 1950 and 2016.

Furthermore, I limited the examination to papers with a minimum of five references. This selection criterion serves dual objectives: firstly, to utilize reference information as a proxy for a paper's knowledge sources, essential for assessing temporal diversity, and secondly, to exclude non-research articles and incomplete records. Subsequently, I narrowed the focus exclusively to journal articles, recognizing the distinctive citing behavior across various scientific document types. Consequently, the dataset comprises a total of 31,266,473 research papers.

The MAG taxonomy provides a structured framework for categorizing research fields, consisting of two hierarchical levels. At the primary level, it encompasses 19 broad research fields, including Medicine, Physics, and Computer Science. The secondary level further delineates 292 subfields. Each paper in the MAG dataset is associated with one or more labels, generated through advanced machine learning models. These labels are accompanied by probability values, indicating the confidence level of the machine-generated predictions. In the analytical framework, I accorded priority to the second-level label for each paper, selecting the label with the highest confidence when multiple labels are present.

4. Methodology

4.1. Temporal diversity

This paper introduces a theoretical and methodological framework for evaluating the temporal diversity of scientific knowledge, which pertains to the diversity of scientific knowledge across temporal dimensions. Specifically, temporal diversity is gauged through an examination of the temporal distribution of references cited in scholarly papers. By examining the temporal distribution of references cited in papers, we can gain insights into the dynamic nature of scientific discourse, the progression of ideas over time, and the interplay between historical foundations and contemporary research (Mukherjee et al., 2017; Zeng et al., 2017).

Drawing inspiration from conceptual frameworks employed in interdisciplinarity and biological diversity (Gates et al., 2019; Stirling, 2007), this study delineates temporal diversity into three primary dimensions: temporal variety, temporal imbalance, and temporal disparity.

For a given paper i and the set of references $R_i = \{r1, r2, \ldots\}$, the ages of the references are mapped into a temporal distribution vector $\{age_{r1}, age_{r2}, \ldots, age_{rk}, \ldots\}$, where age_{rk} represents the age of reference r_k.

(1) Temporal variety is represented by the number of ages covered by the references of paper i, defined as:

$$Temporal\ variety = ln\left(\sum_{r \in R_i} year_r\right) \tag{1}$$

(2) Temporal imbalance is based on the normalized Shannon entropy of the distribution of reference ages, defined as:

$$Temporal\ imbalance = Shannon\ entropy_{age} = -\frac{\sum_{i=1}^{Max(variety)} p_\alpha ln(p_\alpha)}{ln(Max(variety))} \tag{2}$$

where $\alpha \in A_i$ represents an age of the references of paper i, p_α represents the probability of references of that age, and $Max(variety)$ represents the maximum value of temporal variety, which is 178 in this study. Note that I utilized "imbalance" instead of "balance" because it better captures the unevenly distributed nature of different ages of references. When a paper cites only one single year's references, its imbalance will be 0, as will be the case for other temporal diversity measures. Imbalance not only demonstrates a positive relationship with other temporal diversity measures, but also exhibits a similar trend over time to those measures.

(3) Temporal disparity is represented by the maximum age difference among references cited in paper i, defined as:

$$Temporal\ disparity = ln(Max\{|age_{rk} - age_{rl}||age_{rk}, age_{rl} \in A_i\} + 1) \quad (3)$$

where $|age_{rk} - age_{rl}|$ represents the difference between the ages of references *rk* and *rl*, equivalent to the absolute difference in their publication years.

Temporal diversity sheds light on the pace and dynamics of knowledge dissemination and uptake within scientific communities. Papers that cite references spanning a wide range of publication years indicate a synthesis of historical and contemporary literature, reflecting a nuanced understanding of the field's evolution (Bloom et al., 2020; Jones, 2009). As shown in Fig. 1, temporal diversity measures can be applied not only at the reference level but also at the entity or team level (Azoulay et al., 2011; Yang, Xu, et al., 2024). However, this paper's focus is confined to the reference level, and results pertaining to other levels are beyond the scope of this study.

The temporal diversity measures of interest in this study are defined based on the age of reference lists. Accordingly, the values of temporal variety, temporal imbalance, and temporal disparity should correlate with the number of references and the average age of the references. To investigate this, I analyzed the relationship between temporal diversity measures and reference count, as well as the relationship between temporal diversity measures and reference age for all papers in our dataset. Fig. 2 presents these relationships, where the nodes correspond to the average values of papers in each subfield, and the R^2 values indicate the OLS regression without any controls. We also show the relationship between temporal diversity measures and reference age using regression models with control variables and fixed effects in Table S2.

The results demonstrate that temporal diversity measures are positively correlated with both reference count and reference age. Specifically, there is a strong positive correlation between temporal variety and reference count, and between temporal diversity and reference age. However, the R^2 values for the relationship between temporal imbalance and either reference count or reference age are relatively small. This suggests that simple reference count or reference age cannot adequately substitute for the proposed framework of temporal diversity.

4.2. Dependent variables

4.2.1. Citation and highly cited papers

MAG tracks the annual citations received by scholarly papers, and this study employed citation counts as a measure of scientific impact. Citation-based metrics, however, are influenced by various factors, including differing citation dynamics across papers, temporal changes in citation averages, and subfield-specific dependencies (Fortunato et al., 2018). To account for these variations, I used the cumulative number of citations a paper receives five years post-publication (C5). Additionally, in the robustness checks presented in the Supplementary Materials, I also utilized the ten-year citation count (C10) as an alternative impact measure.

Given that citation distributions frequently exhibit a heavy-tail pattern (Wang et al., 2013), this study defined highly cited papers as the top 10 % of highly cited papers (C5) within their respective year and subfield (292 second-level fields in MAG) (Mukherjee et al., 2017). This method also helps to account for the disparities in citation practices across different disciplines and publication years.

$$Highly\ cited\ paper = \begin{cases} 1, if\ top\ 10\%\ C5\ in\ publication\ year\ and\ field \\ 0, otherwise \end{cases} \quad (4)$$

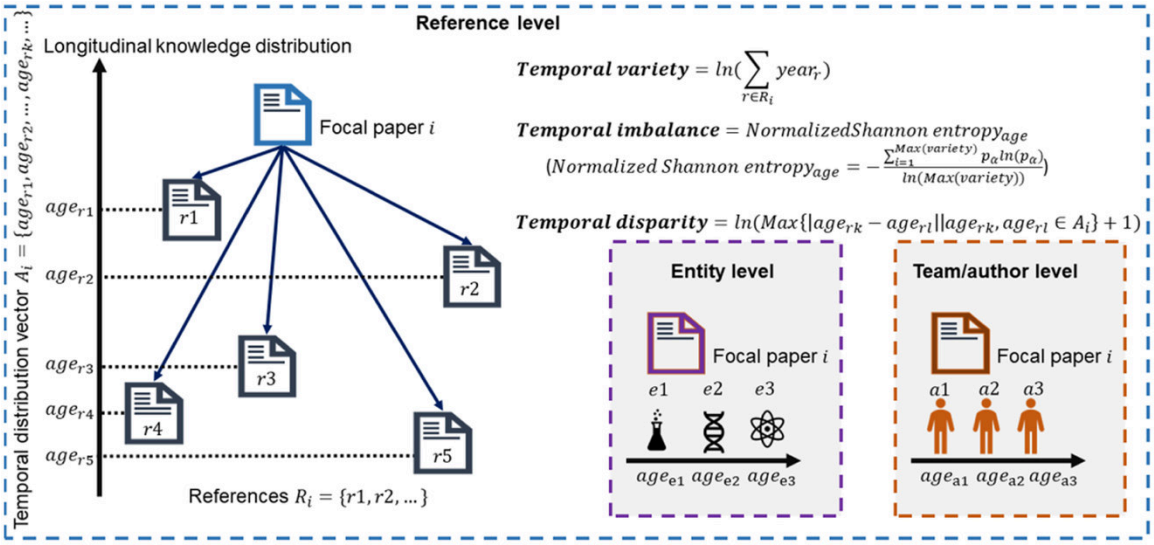


Fig. 1. A framework to quantify the temporal diversity through variety, imbalance and disparity.

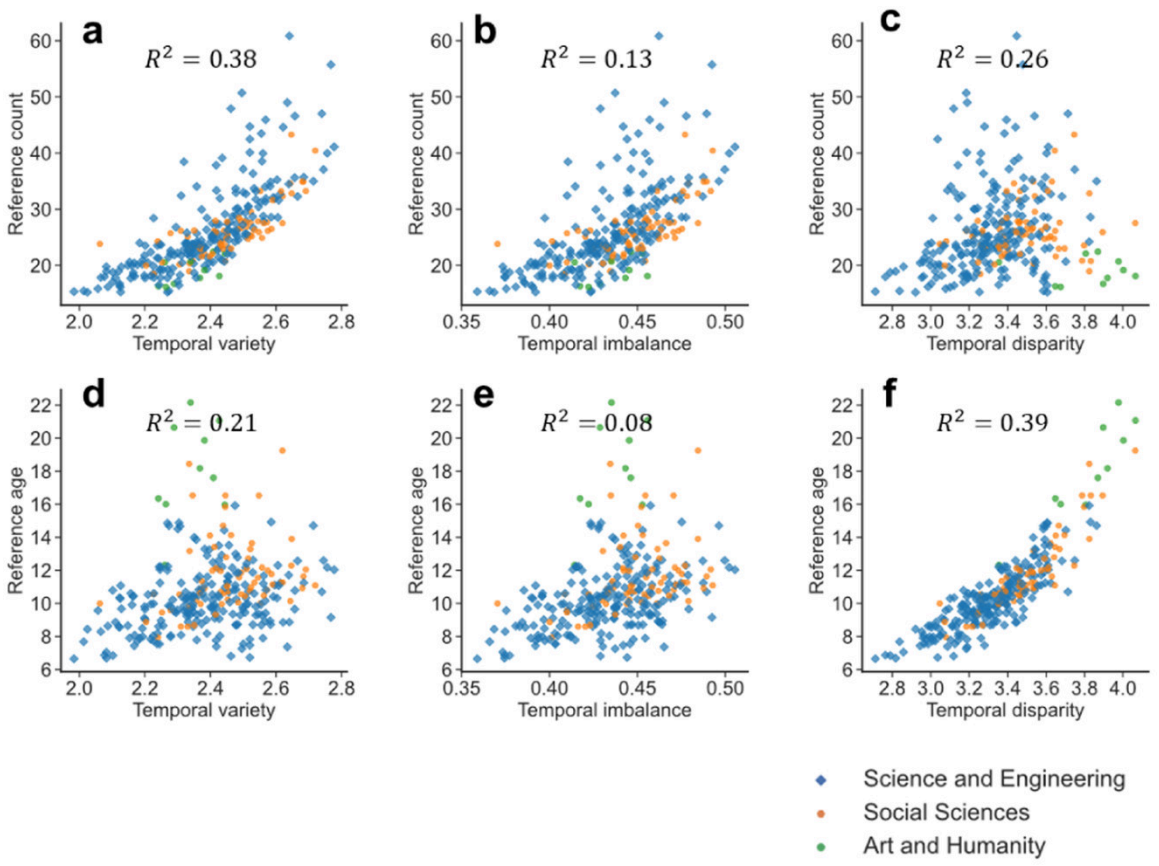


Fig. 2. Temporal diversity measures are positively correlated with reference count as and reference age. The nodes correspond to the average values of papers in each subfield (292 second level fields in MAG). R^2 values indicate the OLS regression without any controls.

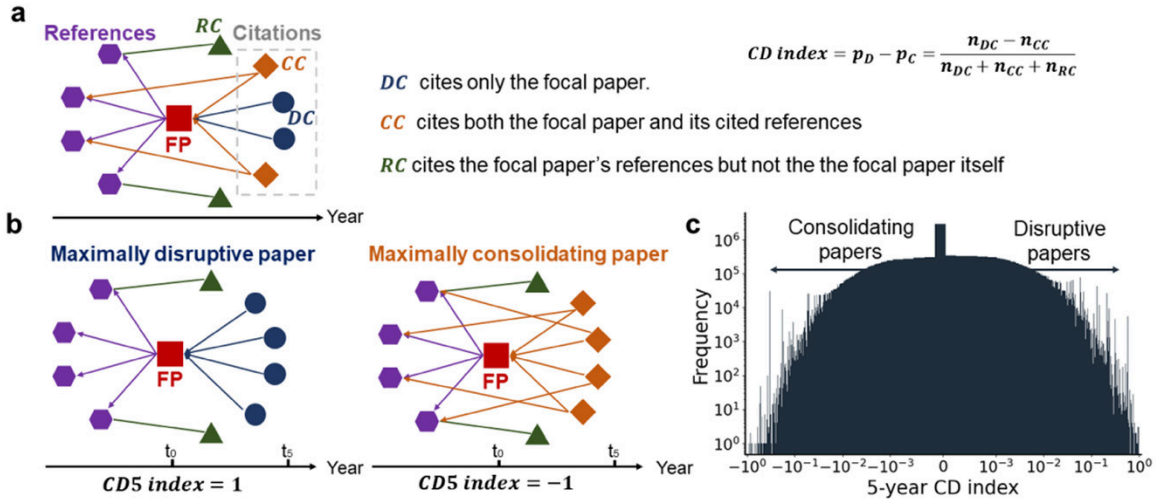


Fig. 3. Quantifying the CD index and disruptive papers. (a) The equation and an illustrative calculation for the CD index. (b) Two hypothetical citation networks, where the CD index is at the maximally disruptive value (CD5 = 1) and maximally consolidating value (CD5 = −1). (c) Distribution of CD5 for papers from MAG (n = 26,730,122) between 1950 and 2016. We consider a paper to be disruptive when its CD5 index is above zero.

4.2.2. CD index and disruptive papers

In addition to citation-based metrics, there exists a growing interest in identifying papers that possess disruptive potential, capable of challenging prevailing scientific paradigms. The CD index, a novel metric extensively validated in literature (Funk & Owen-Smith, 2017; Lin et al., 2023; Park et al., 2023; Wu et al., 2019), quantifies the degree to which an idea disrupts the current state of science or technology. The CD index is defined as:

$$CD\ index = p_D - p_C = \frac{n_{DC} - n_{CC}}{n_{DC} + n_{CC} + n_{RC}} \tag{5}$$

where n_{DC} (the number of disruptive citations) represents the count of publications citing the focal paper without referencing its cited sources, n_{CC} (the number of consolidating citations) denotes the count of publications citing the focal paper and its references, and n_{RC} represents the number of publications citing the focal paper 's references without citing the focal paper itself.

As shown in Fig. 3, a lower CD index suggests extensive citation of references aligning with established knowledge, whereas higher values indicate papers with the potential to induce paradigm shifts (Yang, Deng, et al., 2023; Yang, Gong, et al., 2024; Yang, Hu, et al., 2023). However, the CD index has several limitations in evaluating the disruptive potential of different papers (Bornmann et al., 2020; Leibel & Bornmann, 2023; Yang, Yan, et al., 2024). Primarily, it is a citation-based metric and thus inherits the same limitations as citation counts, such as varying citation dynamics across different papers, temporal changes in the average number of citations (Park et al., 2023), and dependencies on specific subfields (Chu & Evans, 2021). To mitigate these issues, I employed a five-year window to compute the CD index, referred to as CD5.

Moreover, a significant concern with the CD index is its distribution, which is highly centered around zero (Leibel & Bornmann, 2023; Yang, Gong, et al., 2024). To address this, I followed existing studies by setting a dummy variable (Lin et al., 2023), considering a paper to be disruptive when its CD5 index is above zero. This approach allows us to circumvent many limitations of the CD index.

$$Disruptive\ paper = \begin{cases} 1, if\ CD5 > 0 \\ 0, otherwise \end{cases} \tag{6}$$

4.3. Control variables

To address potential sources of bias in the analysis, I integrated a set of fixed effects and control variables into the modeling framework. These variables are designed to account for various factors that may influence measures of temporal diversity, impact, and innovation, thereby enhancing the validity and reliability of the findings.

Firstly, I employed fixed effects for 292 second-level subfields in MAG to control for variations among distinct scientific domains. This approach acknowledges that scientific impact and innovation can vary significantly across disciplines (Chu & Evans, 2021), ensuring a nuanced understanding of the impact of international collaboration on research outcomes.

Secondly, I incorporated year fixed effects spanning from 1950 to 2016 to capture temporal variations and trends. This allows for a comprehensive examination of changes in temporal diversity, impact, and innovation while accounting for potential temporal fluctuations (Park et al., 2023).

Thirdly, journals significantly influence citation and reference patterns due to their diverse scopes, audiences, and editorial policies. Including journal fixed effects helps mitigate these influences and reflects the unique research themes and standards associated with each journal. This study incorporated journal fixed effects to account for heterogeneity related to publication outlets, encompassing a total of 39,893 distinct journals. By integrating these fixed effects, the research aimed to address variations in reference patterns and innovation potential associated with different publication venues and research topics within the analysis (Peng et al., 2021).

Thirdly, journals can significantly impact citation and reference patterns due to their differing scopes, audiences, and editorial policies. Including journal fixed effects helps control for these influences and also reflects the specific research topics and standards associated with each journal. I included journal fixed effects to address heterogeneity related to publication outlets, covering a total of 39,893 distinct journals. By incorporating these fixed effects, this study aimed to account for variations in reference patterns and innovation potential attributed to different publication venues and research topics within the analysis (Peng et al., 2021), thereby enhancing the rigor and validity of the findings.

Furthermore, I recognized the importance of controlling for the number of references cited in each paper. Given that temporal diversity measures are derived from references, variations in reference practices across papers could potentially confound the results (Zeng et al., 2017). Thus, I included the number of references as a control variable in the analysis to mitigate this potential source of bias.

Additionally, it is imperative to account for team-level factors that may influence research outcomes (Lin et al., 2023; Wu et al., 2019). I incorporated several control variables to capture different dimensions of collaboration within research teams: (1) I considered the number of authors involved in a research paper as a critical factor reflecting the collaborative nature of scientific inquiry. (2) I utilized a dummy variable to account for international collaborations, recognizing the global reach and potential impact of such collaborative endeavors (Wagner et al., 2019). (3) To elucidate the knowledge diversity inherent in collaborative research, I evaluated the inter-field diversity of team members as a proxy for the breadth of expertise accessible to the team (Lin et al., 2023). This measure involves classifying the primary field of each team member using the nineteen top-level field-of-study labels in the MAG dataset, thereby distinguishing between monodisciplinary and interdisciplinary teams.

Lastly, to control for the influence of author prestige and ability, I further introduced several author-level covariates, including the

average career age of authors (Yang, Xu, et al., 2024), average number of past publications at the publication year, and average number of top 10 % highly cited papers at the publication year (Brogaard et al., 2024). Additionally, I considered the average rank of affiliation of authors, which is based on the total citation count of all publications affiliated with that institution, to account for the prestige of author affiliations (Hvide & Jones, 2018). Furthermore, in recognition of the potential impact of grant funding on scientific research outcomes (Yang, 2024), this study introduced a control variable to assess whether a paper received grant funding from the National Institutes of Health (NIH) or the National Science Foundation (NSF).

4.4. Empirical specification

To analyze the association between temporal diversity and scientific impact, this study employed a fixed-effect logistic regression model. This model evaluates the effect of temporal diversity on the likelihood of a paper being a highly cited paper, defined as being in the top 10 % of 5-year citations within its publication year and subfield:

$$P(Highly\ cited\ paper_i) = \frac{1}{1+e^{-\left(\alpha+\beta_1\left(Diversity_i\right)+\sum_{k=2}^{K}\beta_k\left(Control_{k,i}\right)+\gamma_t Y_t+\delta_f F_f+\varphi_j J_j\right)}} + \in_i \quad (7)$$

where i represents the article, t the year of publication, f the field, and j the journal. $Diversity_i$ encompasses independent variables such as temporal variety, Temporal imbalance, and temporal disparity. All count variables undergo natural log-transformation within the models to ensure appropriate scaling. The model structure accounts for numerous factors: Y_t denotes year fixed effects, F_f captures second-level field fixed effects, J_j represents journal fixed effects, and $\in_i$ denotes the error term.

Using similar settings, I employed a fixed-effect logistic regression model to analyze the association between temporal diversity and disruption. Specifically, I evaluated the effect of temporal diversity on the likelihood of a paper being disruptive, defined as having a CD5 score greater than 0:

$$P(CD5_i > 0) = \frac{1}{1+e^{-\left(\alpha+\beta_1\left(Diversity_i\right)+\sum_{k=2}^{K}\beta_k\left(Control_{k,i}\right)+\gamma_t Y_t+\delta_f F_f+\varphi_j J_j\right)}} + \in_i \quad (8)$$

where i represents the article, t the year of publication, f the field, and j the journal.

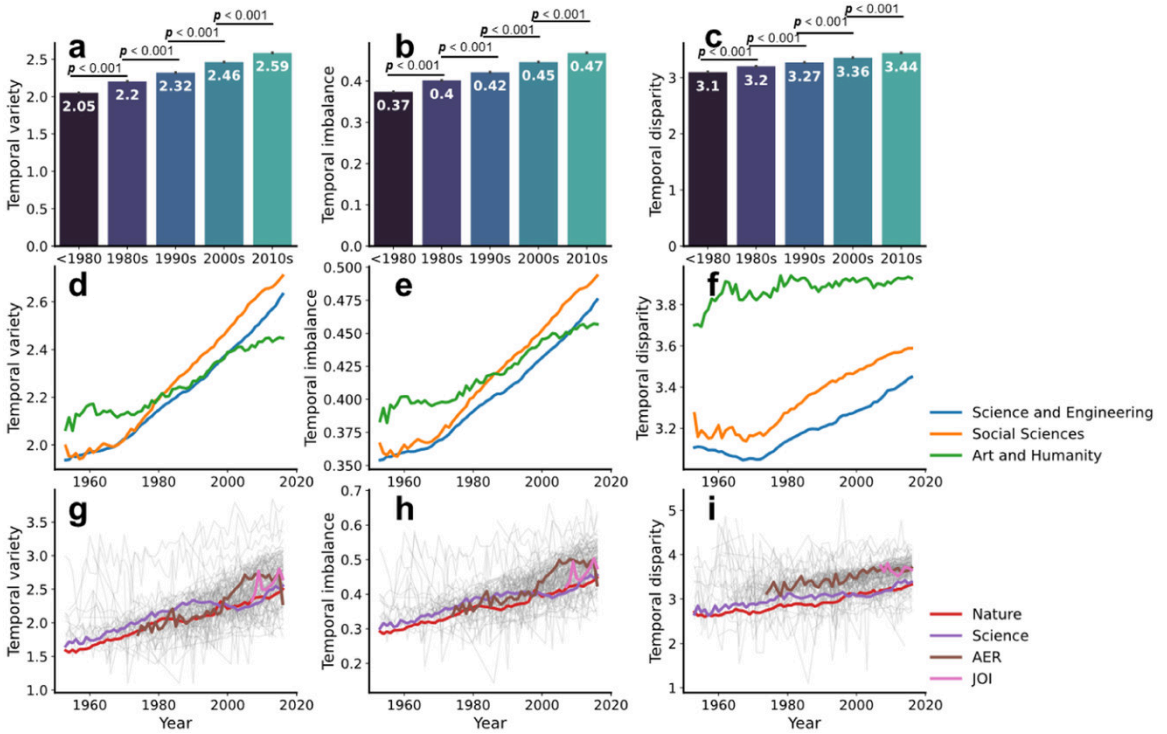


Fig. 4. Temporal diversity increases over time. (a-c) Average temporal diversity (temporal variety, temporal imbalance, temporal disparity) of all papers in each period. Error bars indicate 95 % CI. Two-tailed Welch's *t*-tests are employed to compare diversity in two consecutive periods, all indicating increasing patterns. (d-f) The increasing trend of temporal diversity remains consistent across different fields of study. (g-i) The increasing trend of temporal diversity remains consistent across different journals, with a focus on *Nature, Science, American Economic Review*, and *Journal of Informetrics*.

Furthermore, this study also employed multivariate fixed-effect Ordinary Least Squares (OLS) regression models to analyze the association between temporal diversity and other measures (dependent variables), including atypicality, interdisciplinarity, and delayed citation. I also used OLS to conduct robust analysis based on raw citation count and the CD index:

$$DV_i = \alpha + \beta_1(Diversity_i) + \sum_{k=2}^{K} \beta_k(Control_{k,i}) + \gamma_t Y_t + \delta_f F_f + \varphi_j J_j + \in_i \quad (9)$$

where i represents the article, t the year of publication, f the field, and j the journal.

5. Main results

5.1. The dynamic trend of temporal diversity

Fig. 4 illustrates a discernible increase in temporal variety, imbalance, and disparity emerges as time progresses. Statistical analyses employing Welch's t-tests compare diversity in two consecutive periods: 1950–1979, 1980s, 1990s, 2000s, and 2010s (all p-values $<$ 0.001, see Fig. 4a–c). The dynamic patterns of temporal diversity across different fields further unveils that all fields exhibit increasing trends of temporal variety, imbalance, and disparity since 1950 (Fig. 4d–f). Notably, Art and Humanities papers manifest significantly higher temporal disparity compared to papers in Science and Engineering and Social Sciences. Moreover, Art and Humanities papers demonstrate significantly higher temporal variety and imbalance from 1950 to 1980 but showcase lower temporal variety and imbalance since 2010. These results indicate the change in knowledge patterns across different fields.

An investigation into the dynamic patterns of temporal diversity across different journals is also conducted. I randomly selected 100 journals and plot the yearly temporal diversity values of all papers published in each of them (grey lines in Fig. 4g–i). It is observed that temporal variety, imbalance, and disparity all exhibit increasing patterns across different journals. Four journals—*Nature, Science, American Economic Review* (AER), and *Journal of Informetrics* (JOI)—are highlighted, and they also demonstrate increasing patterns in temporal variety, imbalance, and disparity.

It is important to note that the trend of temporal diversity measures can be influenced by the number of references (Fig. S5). The group with a small reference count (e.g., 5–10 references) demonstrates an almost consistently growing trend over time. In contrast,

Table 1
Estimated effect of temporal diversity on scientific impact and disruption.

Models	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable: Highly cited paper Logistic				Dependent variable: CD5>0 Logistic			
Temporal variety		-1.183***				0.9277***		
		(0.0028)				(0.0018)		
Temporal imbalance			-5.697***				4.501***	
			(0.0121)				(0.0080)	
Temporal disparity				-0.2869***				0.1923***
				(0.0013)				(0.0008)
ln(#Reference)	1.044*	1.692***	1.531***	1.156***	-0.2793*	-0.8290***	-0.7101***	-0.3532***
	(0.0013)	(0.0020)	(0.0017)	(0.0014)	(0.0008)	(0.0013)	(0.0011)	(0.0008)
ln(Team size)	0.4357*	0.4384***	0.4396***	0.4275***	0.0264*	0.0307***	0.0309***	0.0341***
	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0009)	(0.0009)	(0.0009)	(0.0009)
International team	-0.0062	0.0641***	0.0732***	0.0212***	0.1773*	0.1337***	0.1278***	0.1601***
	(0.0022)	(0.0022)	(0.0022)	(0.0022)	(0.0013)	(0.0013)	(0.0013)	(0.0013)
Inter-field team	-0.0189*	-0.0178***	-0.0180***	-0.0160***	-0.0640*	-0.0653***	-0.0653***	-0.0663***
	(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0013)	(0.0013)	(0.0013)	(0.0013)
Grant funding	-0.0752*	-0.0741***	-0.0738***	-0.0740***	0.0853*	0.0851***	0.0850***	0.0848***
	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0011)	(0.0011)	(0.0011)	(0.0011)
ln(Career age+1)	-0.0181*	0.0019	0.0065**	-0.0151***	-0.0501*	-0.0575***	-0.0603***	-0.0506***
	(0.0022)	(0.0022)	(0.0022)	(0.0022)	(0.0016)	(0.0017)	(0.0017)	(0.0016)
ln(#Past pubs+1)	-0.3427*	-0.3610***	-0.3632***	-0.3513***	-0.1524*	-0.1423***	-0.1411***	-0.1468***
	(0.0016)	(0.0016)	(0.0016)	(0.0016)	(0.0009)	(0.0009)	(0.0009)	(0.0009)
ln(#Past highly citeds+1)	0.8403*	0.8123***	0.8066***	0.8331***	-0.0369*	-0.0096***	-0.0045***	-0.0305***
	(0.0013)	(0.0014)	(0.0014)	(0.0013)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
ln(Affiliation Rank)	0.0105*	0.0112***	0.0113***	0.0104***	0.0095*	0.0089***	0.0088***	0.0095***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	29,370,484	29,370,484	29,370,484	29,370,484	26,504,315	26,504,315	26,504,315	26,504,315
Pseudo R2	0.26875	0.27831	0.2804	0.27126	0.08172	0.08934	0.09122	0.08347

Note: Heteroskedasticity-robust standard errors are reported in parentheses. *p $<$ 0.05; ** p $<$ 0.01; *** p $<$ 0.001. Logistic regression is used to estimate the effect of temporal diversity on the probability of being the top 10 % highly cited paper or disruptive paper. The odds ratios are not reported but can be calculated as e^{coef}.

the group with a large reference count shows a decreasing trend during the earlier period (1950–2000). It is important to note that papers from earlier periods typically have a smaller average reference count, with most containing only a small number of references. And after 2000, there is a consistently growing trend regardless of reference count. Therefore, the overarching trend is that all temporal diversity measures increase over time.

5.2. *Temporal diversity is associated with lower impact and higher disruption*

Table 1 reports the effect of temporal diversity on the likelihood of a paper being a highly cited or a disruptive one based on the fixed-effect logistic regression models, while controlling for confounding factors including the year fixed effects, subfield fixed effects, journal fixed effects, reference count, team-level controls, and author-level factors.

Models (1–4) consistently demonstrate a significantly negative association between temporal diversity and scientific impact. Specifically, a one-S.T.D. (0.525) increase in temporal variety leads to the logit odds of highly cited papers occurring decreasing by 0.463 ($e^{-1.183*0.525}$-1) on average. Similarly, a one-S.T.D. (0.095) increase in Temporal imbalance results in the logit odds of highly cited papers occurring decreasing by 0.418 ($e^{-5.697*0.095}$-1) on average. Moreover, a one-S.T.D. (0.667) increase in temporal disparity leads to the logit odds of highly cited papers occurring decreasing by 0.174 ($e^{-0.2869*0.667}$-1) on average.

On the other hand, Models (5–8) consistently demonstrate a significantly positive effect of temporal diversity on disruption potentials. Specifically, a one-S.T.D. (0.525) increase in temporal variety leads to the logit odds of disruptive papers (CD5>0) occurring increasing by 0.628 ($e^{0.9277*0.525}$-1) on average. Similarly, a one-S.T.D. (0.095) increase in Temporal imbalance results in the logit odds of disruptive papers occurring increasing by 0.534 ($e^{4.501*0.095}$-1) on average. Moreover, a one-S.T.D. (0.667) increase in temporal disparity leads to the logit odds of disruptive papers occurring increasing by 0.137 ($e^{0.1923*0.667}$-1) on average.

The regression findings indicate that temporal diversity exhibits a negative association with citation impact (highly cited papers) while showing a positive association with disruptive potentials (CD5 > 0). These associations are consistently significant across various

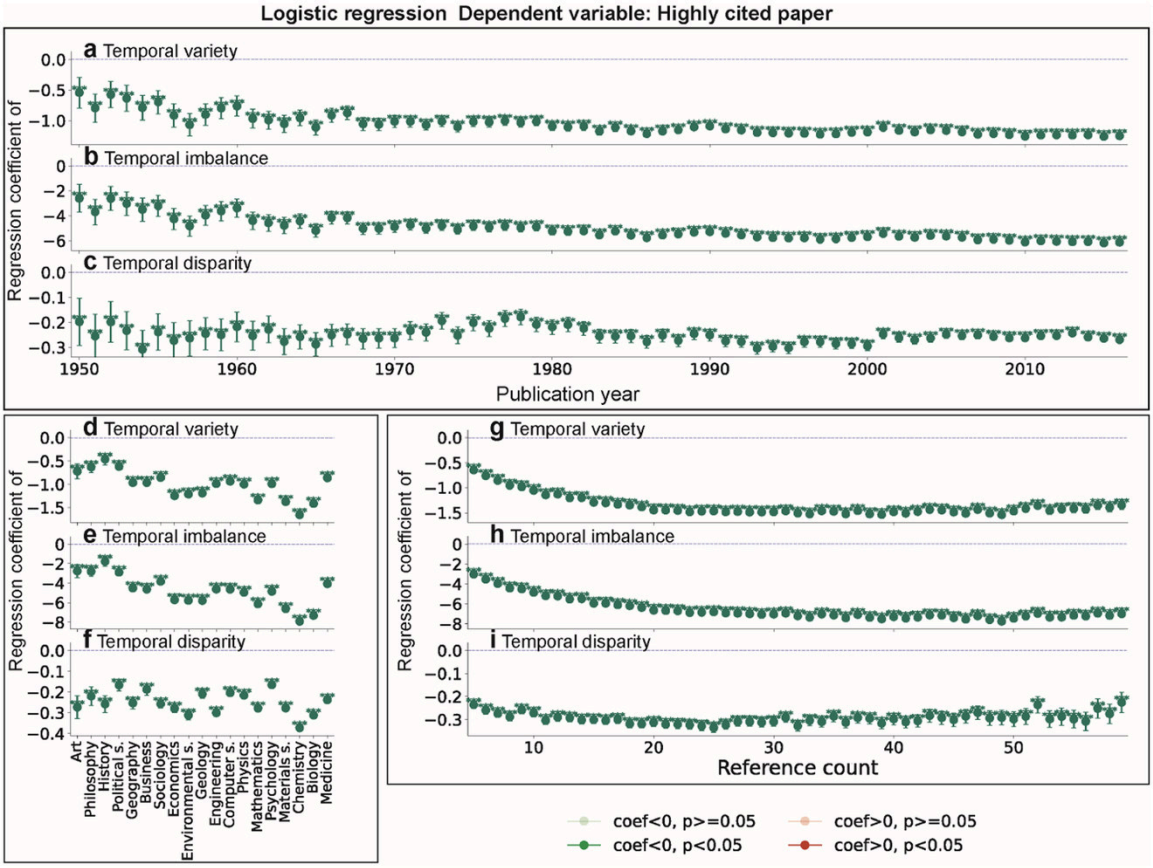


Fig. 5. Split sample regression: Logistic estimated effect of temporal diversity on the top 10 % highly cited papers. (a-c) Estimated effect of temporal diversity (temporal variety, temporal imbalance, temporal disparity) on highly cited papers in each year, spanning from 1950 to 2016. (d-f) Estimated effect of temporal diversity on highly cited papers across 19 fields. (g-i) Estimated effect of temporal diversity on highly cited papers for different reference counts (ranging from 5 to 60). All control variables and fixed effects are included in the logistic regression. Error bars depict 95 % CI based on robust standard errors. *p < 0.05; ** p < 0.01; *** p < 0.001.

dimensions of temporal diversity, including temporal variety, temporal imbalance, and temporal disparity.

To ensure the robustness of the results, I conducted regression analyses omitting some controls, and these associations remained consistently significant. Furthermore, I performed OLS regressions using raw 5-year citation counts and the 5-year CD index, as well as robustness analyses using 10-year citation counts and the 10-year CD index (from 1950 to 2011). The results presented in Tables S3 and S4 confirm that temporal diversity is negatively associated with citation impact and positively associated with disruptive potentials, further validating the robustness of these associations.

Furthermore, I conducted a set of split-sample regressions to examine the effect of temporal diversity on highly cited papers across various contextual dimensions, including different years, fields, and reference counts, given the substantial sample size in the analysis. The results are presented in Fig. 5, which illustrates the logistic regression coefficients of temporal diversity within each setting while controlling for all other confounding factors. Each node in the figure displays the regression coefficient of temporal diversity (temporal variety, temporal imbalance, temporal disparity).

Remarkably, all estimated coefficients of temporal diversity are consistently negative (all p values < 0.001), indicating that the negative association between temporal diversity (temporal variety, temporal imbalance, temporal disparity) and citation impact holds across different years, fields, and reference count. This finding also suggests that variations in time, discipline, and reference count cannot explain the negative association between temporal diversity and citation impact.

Similarly, I conducted a set of split-sample regressions to assess the effect of temporal diversity on disruptive papers (CD5 > 0) across different years, fields, and reference counts. The results are depicted in Fig. 6, illustrating the logistic regression coefficients of temporal diversity within each setting while controlling for all other confounding factors.

All estimated coefficients of temporal diversity are consistently positive (all p values < 0.001), indicating that the positive association between temporal diversity and disruptive potentials holds across different years, fields, and reference counts. This finding suggests that variations in time, discipline, and reference count cannot explain the positive association between temporal diversity and disruption.

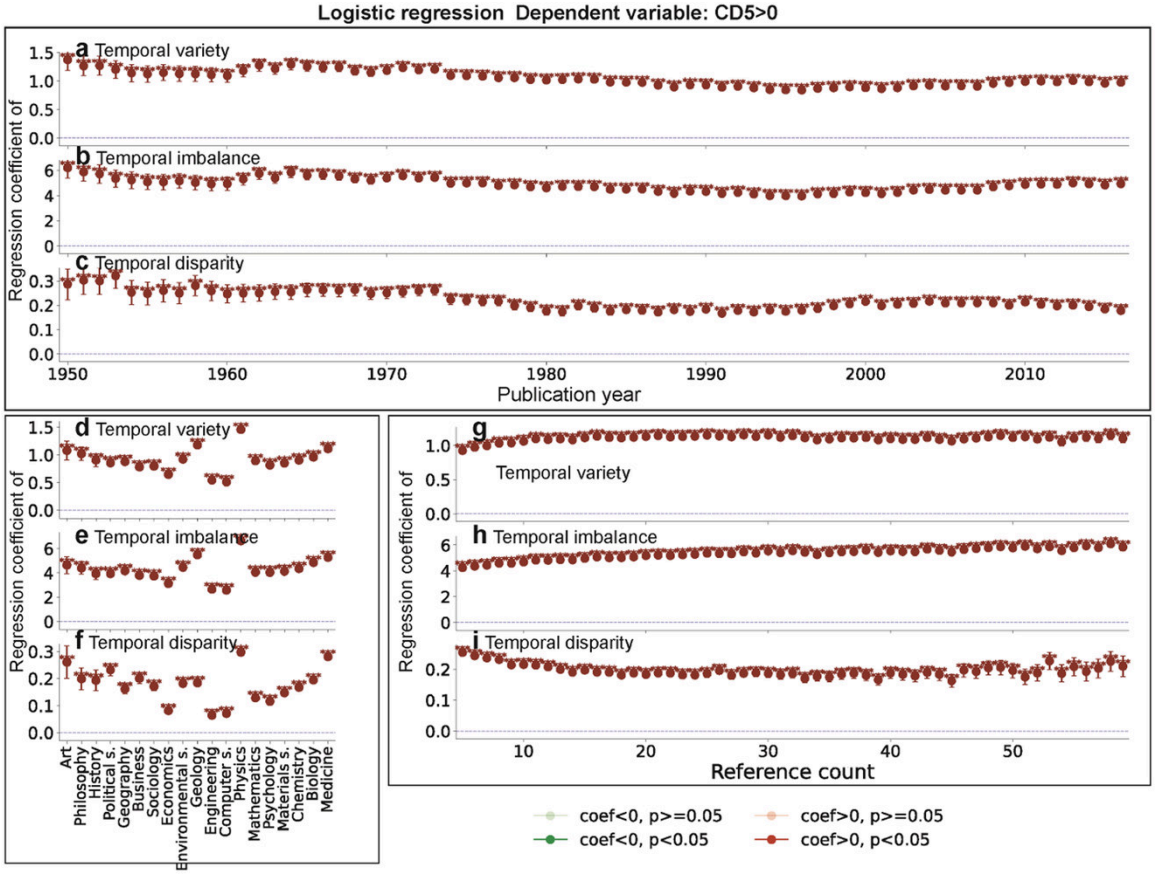


Fig. 6. Split sample regression: Logistic estimated effect of temporal diversity on disruptive papers (CD5>0). (a-c) Estimated effect of temporal diversity (temporal variety, temporal imbalance, temporal disparity) on disruptive papers in each year, spanning from 1950 to 2016. (d-f) Estimated effect of temporal diversity on disruptive papers across 19 fields. (g-i) Estimated effect of temporal diversity on disruptive papers for different reference counts (ranging from 5 to 60). All control variables and fixed effects are included in the logistic regression. Error bars depict 95 % CI based on robust standard errors. *p < 0.05; ** p < 0.01; *** p < 0.001.

5.3. Controlling for reference age

In the previous analysis, the average reference age was not included in the regression model. This omission was intentional to provide intuitive and consistent initial results. However, it is crucial to recognize that all temporal diversity measures are positively correlated with reference age (see Fig. 2 and Tables S2). To fully comprehend the role of temporal diversity in scientific impact and disruption, it is essential to analyze the results while controlling for reference age.

In this subsection, the analysis was conducted with the average reference age controlled. The results, as reported in Table 2, indicate that even when controlling for average reference age, the effects of temporal variety and temporal imbalance on scientific impact remain negatively significant. Conversely, the effect of temporal disparity on scientific impact becomes positively significant. This positive significance can be attributed to the high positive correlation between temporal disparity and reference age (see Fig. 2).

Regarding disruption, the analysis reveals that, after controlling for reference age, the effects of all temporal diversity measures become negatively significant. This contrasts with the previously observed positively significant effects when reference age was not controlled. This shift indicates that the age of referenced knowledge is a primary factor contributing to higher disruption. When average reference age is controlled, temporal diversity alone appears to lead to lower disruption.

I further conducted a series of split-sample regressions to evaluate the effect of temporal diversity on the top 10 % highly cited papers and disruptive papers ($CD5 > 0$) across different groups categorized by average reference age (binned in increments of 5 years). The results are illustrated in Fig. 7, which displays the logistic regression coefficients for temporal diversity within each category while controlling for all other confounding factors.

Notably, the negative coefficients for temporal variety and temporal imbalance concerning highly cited papers remain consistent across the different reference age groups. In contrast, the coefficient for temporal disparity on highly cited papers is negative only for the lowest reference age group and becomes positive for the larger reference age groups. Regarding disruption, the coefficients for temporal variety, temporal imbalance, and temporal disparity are all positive for the smallest reference age group but shift to negative for the larger reference age groups. Interestingly, in the largest reference age group, the coefficients for temporal variety and temporal imbalance revert to positive. These findings suggest that reference age may mediate the effect of temporal diversity on disruption.

5.4. Other possible mechanisms

Besides the reference age, one possible pathway by which temporal diversity results in a lower citation count while increasing disruption is through the generation of more interdisciplinary and atypical combinations of knowledge. The proposed framework offers a longitudinal perspective on the temporal distribution of knowledge. I measured interdisciplinarity and atypical knowledge combinations using distributions within reference disciplines, reflecting disciplinary knowledge diversity (Fontana et al., 2020).

Table 3 presents the relationships between temporal diversity and interdisciplinarity (measured by the Rao-Stirling index), as well as between temporal diversity and atypical knowledge combinations (measured by atypicality), using fixed-effect OLS regression models. These models account for all confounding factors. The results consistently show a positive association between temporal diversity (including temporal variety, temporal imbalance, and temporal disparity) and both interdisciplinarity and atypical combinations of knowledge. This indicates that greater diversity in knowledge over time correlates with increased diversity in disciplinary knowledge.

Another possible pathway by which temporal diversity results in a lower citation count while increasing disruption is that temporally diverse knowledge may lead to delayed citations. To investigate this, I employed multivariate fixed-effect OLS regression models to explore the impact of temporal diversity on delayed citation ($\boldsymbol{t_\alpha}$). As demonstrated in Table 3, models 7–9 consistently reveal a positive association between temporal diversity and delayed citation. This suggests that temporal diversity indeed contributes to delayed citation in scientific research.

Subsequently, I incorporated potential pathway variables, including interdisciplinarity (measured by the Rao-Stirling index), atypical knowledge combinations (atypicality), delayed citation, and reference age, into the logistic regression models, while controlling for other confounding factors.

Fig. 8 demonstrates that these pathway variables do not substantially alter the direction of the coefficients, with the exception of reference age. This indicates that interdisciplinarity, atypical knowledge combinations, and delayed citation cannot explain the negative association between temporal diversity and citation impact, nor do they account for the positive relationship between temporal diversity and disruptive potential.

On the other hand, the findings suggest that higher temporal diversity leading to greater disruption may be primarily due to the use of older references. However, the disadvantages of temporal diversity in terms of citation impact cannot be entirely explained by these factors alone. This underscores that the temporal diversity of knowledge is an underestimated yet significant correlate of lower citation impact.

6. Discussion

6.1. Implications

This paper contributes to the field of science of science and innovation by advancing our understanding of temporal diversity in scientific knowledge and its implications for scientific impact and innovation. Firstly, the study introduces a novel framework for assessing temporal diversity in scholarly research. By quantifying temporal variety, imbalance, and disparity across a comprehensive

Table 2
Estimated effect of temporal diversity on scientific impact and disruption while controlling for the average reference age.

Models	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable: Highly cited paper Logistic				Dependent variable: CD5>0 Logistic			
Temporal variety		-0.2592*** (0.0041)				-0.7958*** (0.0028)		
Temporal imbalance			-2.283*** (0.0184)				-2.612*** (0.0122)	
Temporal disparity				0.2222*** (0.0016)				-0.6510*** (0.0012)
ln(Reference age+1)	-0.9493*** (0.0018)	-0.8240*** (0.0027)	-0.6932*** (0.0027)	-1.134*** (0.0022)	1.121*** (0.0012)	1.466*** (0.0018)	1.383*** (0.0018)	1.732*** (0.0017)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	29,370,484	29,370,484	29,370,484	29,370,484	29,370,484	29,370,484	29,370,484	29,370,484
Pseudo R2	0.28138	0.28159	0.28219	0.28236	0.11192	0.11456	0.11341	0.12221

Note: Heteroskedasticity-robust standard errors are reported in parentheses. *$p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

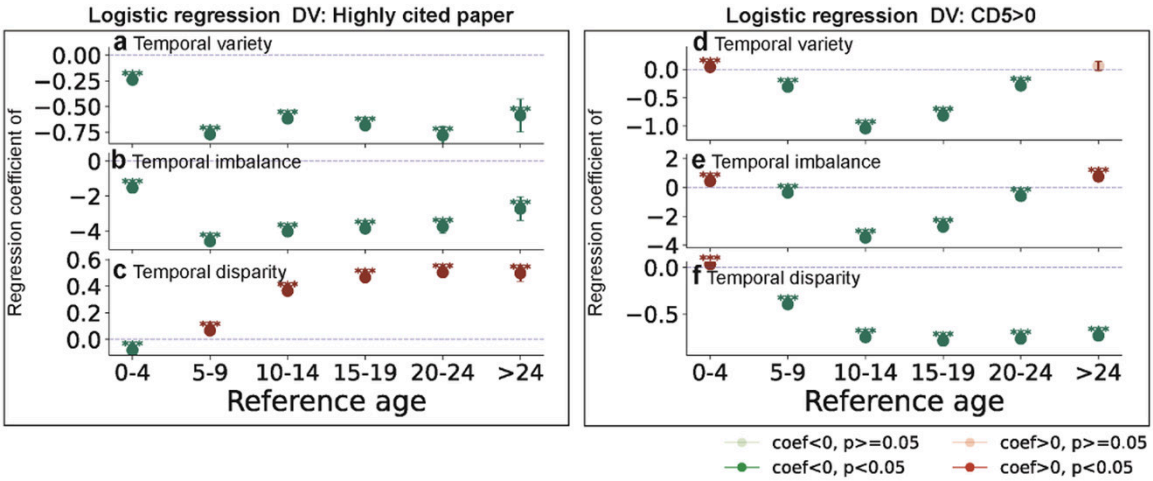


Fig. 7. Split sample regression based on the reference age group. (a-c) Estimated effect of temporal diversity (temporal variety, temporal imbalance, temporal disparity) on the top 10 % highly cited papers in each reference group. (d-f) Estimated effect of temporal diversity (temporal variety, temporal imbalance, temporal disparity) on or disruptive papers (CD5>0) in each reference group. All control variables and fixed effects are included in the logistic regression. Error bars depict 95 % CI based on robust standard errors. *p < 0.05; ** p < 0.01; *** p < 0.001.

dataset of 31 million articles spanning seven decades, this framework enhances our ability to capture the temporal dynamics of scientific knowledge. This contribution is aligned with theoretical discussions on the evolution and dissemination of knowledge over time (Mukherjee et al., 2017). The identification of increasing trends in temporal diversity suggests that scientific knowledge increasingly incorporates insights from both historical and contemporary literature, reflecting a cumulative process of knowledge creation.

Secondly, the study finds a consistent negative association between temporal diversity and citation impact, in line with traditional assumptions about the relationship between the recency of references and academic impact (Herrmannova et al., 2018; Onodera & Yoshikane, 2015). Conversely, the positive association between temporal diversity and disruptive potential underscores the role of temporal synthesis in driving innovation and scientific breakthroughs (Funk & Owen-Smith, 2017; Kuhn, 1962; Park et al., 2023). This finding supports the view that knowledge integration across temporal epochs can lead to novel and potentially transformative insights that challenge established paradigms. This contribution enriches the literature on disruptive innovation in science by providing empirical evidence linking the temporal diversity of references to disruptive potential. Moreover, it enhances theories of knowledge development by suggesting that while recent knowledge may lead to higher citation rates in the short term, papers that synthesize a broader range of temporal perspectives may exhibit higher disruptive potential over time. This highlights the nuanced relationship between the temporal diversity of references and the long-term impact trajectories of scientific papers.

Thirdly, the study identifies that the observed positive association between temporal diversity and disruption can be primarily attributed to the use of older references. By incorporating historical perspectives and foundational ideas, papers with higher temporal diversity may challenge existing paradigms and foster novel insights, thus contributing to disruptive innovation. This finding underscores the value of historical context in shaping scientific breakthroughs and innovative advancements. However, the study also reveals that the disadvantages of temporal diversity in terms of citation impact cannot be fully explained by the use of older references alone. While older references contribute to disruption, they may also lead to delayed recognition and lower citation rates. This suggests that temporal diversity represents an underestimated yet potent correlate of lower citation impact results beyond the mechanisms old knowledge.

Researchers can leverage the insights from this study to enhance the visibility and impact of their work. While recent knowledge increases citation impact, historical knowledge may increase the potential to contribute to disruptive innovations. Policymakers and funding agencies should also consider the trade-offs between disruptive potential and citation impact when assessing the value of research outputs.

6.2. Limitations and outlook

While this study offers valuable insights into the dynamics of temporal diversity in scholarly research, several limitations warrant acknowledgment and provide avenues for future research. First, the metrics used to quantify temporal diversity may have inherent limitations. For example, disciplinary classification schemes and temporal distribution metrics may vary in their granularity and accuracy, potentially influencing the results. Future research could explore alternative approaches to defining temporal diversity, such as utilizing the Rao-Stirling index (Stirling, 2007) and DIV* index (Leydesdorff, 2018), to enhance the accuracy and reliability of diversity measurements.

Second, the study's findings suggest that temporal diversity is negatively associated with citation impact but positively associated with disruption potential. However, the implications of these findings for different scientific disciplines and subfields remain underexplored, and the mechanisms underlying these associations are not yet fully understood. Therefore, caution is warranted in

Table 3
Estimated effect of temporal diversity on atypicality, interdisciplinarity and delayed citation.

Models	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Dependent variable: Rao-Stirling OLS			Dependent variable: Atypicality OLS			Dependent variable: $\ln(t_{\alpha}+1)$ OLS		
Temporal variety	0.1677*** (4.54e-5)			1.059*** (0.0018)			0.2838*** (0.0006)		
Temporal imbalance		0.7314*** (0.0002)			4.240*** (0.0080)			1.278*** (0.0028)	
Temporal disparity			0.0528*** (2.15e-5)			0.5694*** (0.0008)			0.0960*** (0.0003)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	29,703,348	29,703,348	29,703,348	30,106,249	30,106,249	30,106,249	27,928,471	27,928,471	27,928,471
R2	0.4887	0.49225	0.3541	0.26106	0.25976	0.2666	0.15385	0.15432	0.1517

Note: Heteroskedasticity-robust standard errors are reported in parentheses. *$p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

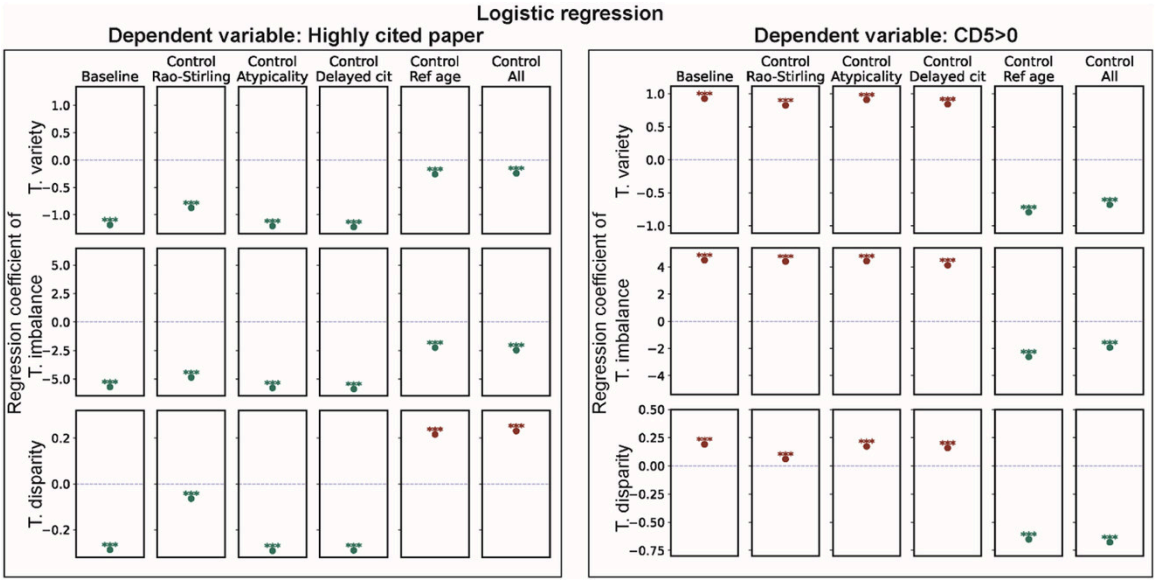


Fig. 8. Mechanism analysis: the association between temporal diversity and scientific impact, as well as the association between temporal diversity and disruption controlling for potential pathways.

interpreting the implications driven by these associations. Future research could investigate disciplinary variations in the relationship between temporal diversity and scientific impact to provide more nuanced insights.

Third, this study predominantly employs retrospective observational data, which limits the ability to establish causal relationships definitively. Experimental studies or quasi-experimental settings could provide stronger evidence on the impact of temporal diversity on scientific innovation. Future research could also delve deeper into the mechanisms driving the observed associations, considering factors such as citation behavior and knowledge diffusion dynamics.

Data availability

The Microsoft Academic Graph data can be downloaded via https://zenodo.org/record/2628216#.Y-7RR_5Bz-g. Other data used in this study can be obtained by making reasonable requests.

CRediT authorship contribution statement

Alex J. Yang: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Conflict of interest

The author declares no competing interest relevant to this article.

Acknowledgments

The author acknowledges the support provided by the Open Fund for Innovative Evaluation from Fudan University (CXPJ2024004). Special thanks are extended to Dr. Yian Yin from Cornell University for invaluable discussions. The author also expresses sincere gratitude to the editor and anonymous reviewers for their insightful feedback, which has significantly enhanced this work.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.joi.2024.101594.

References

Ahlgren, P., Colliander, C., & Sjögårde, P. (2018). Exploring the relation between referencing practices and citation impact: A large-scale study based on Web of Science data. *Journal of the Association for Information Science and Technology, 69*(5), 728–743. https://doi.org/10.1002/asi.23986

Azoulay, P., Graff Zivin, J. S., & Manso, G. (2011). Incentives and creativity: Evidence from the academic life sciences. *The RAND Journal of Economics, 42*(3), 527–554. https://doi.org/10.1111/j.1756-2171.2011.00140.x

Bloom, N., Jones, C. I., Van Reenen, J., & Webb, M. (2020). Are ideas getting harder to find? *American Economic Review, 110*(4), 1104–1144. https://doi.org/10.1257/aer.20180338

Bornmann, L., Devarakonda, S., Tekles, A., & Chacko, G. (2020). Are disruption index indicators convergently valid? The comparison of several indicator variants with assessments by peers. *Quantitative Science Studies, 1*(3), 1242–1259. https://doi.org/10.1162/qss_a_00068

Brogaard, J., Engelberg, J. E., Eswar, S. K., & Wesep, E. D. V. (2024). On the causal effect of fame on citations. *Management Science*. https://doi.org/10.1287/mnsc.2022.00840

Chu, J. S. G., & Evans, J. A. (2021). Slowed canonical progress in large fields of science. *Proceedings of the National Academy of Sciences, 118*(41), Article e2021636118. https://doi.org/10.1073/pnas.2021636118

Fontana, M., Iori, M., Montobbio, F., & Sinatra, R. (2020). New and atypical combinations: An assessment of novelty and interdisciplinarity. *Research Policy, 49*(7), Article 104063. https://doi.org/10.1016/j.respol.2020.104063

Fortunato, S., Bergstrom, C. T., Boerner, K., Evans, J. A., Helbing, D., Milojevic, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., & Barabasi, A.-L. (2018). Science of science. *Science (New York, N.Y.), 359*(6379), eaao0185. https://doi.org/10.1126/science.aao0185

Funk, R. J., & Owen-Smith, J. (2017). A dynamic network measure of technological change. *Management Science, 63*(3), 791–817. https://doi.org/10.1287/mnsc.2015.2366

Furman, J. L., & Stern, S. (2011). Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *American Economic Review, 101*(5), 1933–1963. https://doi.org/10.1257/aer.101.5.1933

Gates, A. J., Ke, Q., Varol, O., & Barabási, A. L. (2019). Nature's reach: Narrow work has broad impact. *Nature, 575*(7781), 32–34. https://doi.org/10.1038/d41586-019-03308-7

Haslam, N., Ban, L., Kaufmann, L., Loughnan, S., Peters, K., Whelan, J., & Wilson, S. (2008). What makes an article influential? Predicting impact in social and personality psychology. *Scientometrics, 76*(1), 169–185. https://doi.org/10.1007/s11192-007-1892-8

Herrmannova, D., Patton, R. M., Knoth, P., & Stahl, C. G. (2018). Do citations and readership identify seminal publications? *Scientometrics, 115*(1), 239–262. https://doi.org/10.1007/s11192-018-2669-y

Hvide, H. K., & Jones, B. F. (2018). University innovation and the professor's privilege. *American Economic Review, 108*(7), 1860–1898. https://doi.org/10.1257/aer.20160284

Jones, B. F. (2009). The burden of knowledge and the "Death of the renaissance man": Is innovation getting harder? *The Review of Economic Studies, 76*(1), 283–317. https://doi.org/10.1111/j.1467-937X.2008.00531.x

Jones, B. F., & Weinberg, B. A. (2011). Age dynamics in scientific creativity. *Proceedings of the National Academy of Sciences, 108*(47), 18910–18914. https://doi.org/10.1073/pnas.1102895108

Klüppel, L., & Knott, A. M. (2023). Are ideas being fished out? *Research Policy, 52*(2), Article 104665. https://doi.org/10.1016/j.respol.2022.104665

Kuhn, T. S. (1962). Historical structure of scientific discovery. *Science (New York, N.Y.), 136*(3518), 760–764.

Kwon, S., Youtie, J., & Porter, A. L. (2020). Interdisciplinary knowledge combinations and emerging technological topics: Implications for reducing uncertainties in research evaluation. *Research Evaluation, 30*(1), 127–140. https://doi.org/10.1093/reseval/rvaa029

Leibel, C., & Bornmann, L. (2023). What do we know about the disruption index in scientometrics? An overview of the literature. *Scientometrics*. https://doi.org/10.1007/s11192-023-04873-5

Leydesdorff, L. (2018). Diversity and interdisciplinarity: How can one distinguish and recombine disparity, variety, and balance? *Scientometrics, 116*(3), 2113–2121. https://doi.org/10.1007/s11192-018-2810-y

Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2019). Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient. *Journal of Informetrics, 13*(1), 255–269. https://doi.org/10.1016/j.joi.2018.12.006

Li, B., Chen, S., & Larivière, V. (2023). Interdisciplinarity affects the technological impact of scientific research. *Scientometrics, 128*(12), 6527–6559. https://doi.org/10.1007/s11192-023-04846-8

Lin, Y., Evans, J. A., & Wu, L. (2022). New directions in science emerge from disconnection and discord. *Journal of Informetrics, 16*(1), Article 101234. https://doi.org/10.1016/j.joi.2021.101234

Lin, Y., Frey, C. B., & Wu, L. (2023). Remote collaboration fuses fewer breakthrough ideas. *Nature, 623*(7989), 987–991. https://doi.org/10.1038/s41586-023-06767-1

Liu, J., Chen, H., Liu, Z., Bu, Y., & Gu, W. (2022). Non-linearity between referencing behavior and citation impact: A large-scale, discipline-level analysis. *Journal of Informetrics, 16*(3), Article 101318. https://doi.org/10.1016/j.joi.2022.101318

Liu, L., Jones, B. F., Uzzi, B., & Wang, D. (2023). Data, measurement and empirical methods in the science of science. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-023-01562-4

Mammola, S., Piano, E., Doretto, A., Caprio, E., & Chamberlain, D. (2022). Measuring the influence of non-scientific features on citations. *Scientometrics, 127*(7), 4123–4137. https://doi.org/10.1007/s11192-022-04421-7

Merton, R. K. (1961). Singletons and multiples in scientific discovery. *Proceedings of the American Philosophical Society, 105*(5), 470. -+.

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago, IL, USA: University of Chicago Press.

Mukherjee, S., Romero, D. M., Jones, B., & Uzzi, B. (2017). The nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology: The hotspot. *Science Advances, 3*(4). https://doi.org/10.1126/sciadv.1601315

Onodera, N., & Yoshikane, F. (2015). Factors affecting citation rates of research articles. *Journal of the Association for Information Science and Technology, 66*(4), 739–764. https://doi.org/10.1002/asi.23209

Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature, 613*(7942), 138–144. https://doi.org/10.1038/s41586-022-05543-x

Peng, H., Ke, Q., Budak, C., Romero, D. M., & Ahn, Y.-Y. (2021). Neural embeddings of scholarly periodicals reveal complex disciplinary organizations. *Science Advances, 7*(17), eabb9004. https://doi.org/10.1126/sciadv.abb9004

Song, Y., Fang, Z., & Fu, H.-Z. (2023). Influence of interdisciplinarity on scientific impact: The case of climate change field. *Journal of Informetrics, 17*(3), Article 101410. https://doi.org/10.1016/j.joi.2023.101410

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of The Royal Society Interface, 4*(15), 707–719. https://doi.org/10.1098/rsif.2007.0213

Sun, Y., Livan, G., Ma, A., & Latora, V. (2021). Interdisciplinary researchers attain better long-term funding performance. *Communications Physics, 4*(1), 263. https://doi.org/10.1038/s42005-021-00769-z

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science (New York, N.Y.), 342*(6157), 468–472. https://doi.org/10.1126/science.1240474

Vantard, M., Galland, C., & Knoop, M. (2023). Interdisciplinary research: Motivations and challenges for researcher careers. *Quantitative Science Studies*, 1–19. https://doi.org/10.1162/qss_a_00265

Wagner, C. S., Whetsell, T. A., & Mukherjee, S. (2019). International research collaboration: Novelty, conventionality, and atypicality in knowledge recombination. *Research Policy, 48*(5), 1260–1270. https://doi.org/10.1016/j.respol.2019.01.002

Wang, D., Song, C., & Barabási, A.-L. (2013). Quantifying long-term scientific impact. *Science (New York, N.Y.), 342*(6154), 127–132. https://doi.org/10.1126/science.1237825

Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics, 94*(3), 851–872. https://doi.org/10.1007/s11192-012-0775-9

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Dong, Y., & Kanakia, A. (2020). Microsoft academic graph: When experts are not enough. *Quantitative Science Studies, 1*(1), 396–413. https://doi.org/10.1162/qss_a_00021

Wu, L. F., Wang, D. S., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature, 566*(7744), 378. https://doi.org/10.1038/s41586-019-0941-9. -+.

Yang, A. J. (2024). Unveiling the impact and dual innovation of funded research. *Journal of Informetrics, 18*(1), Article 101480. https://doi.org/10.1016/j.joi.2023.101480

Yang, A. J., & Deng, S. (2024). Dynamic patterns of the disruptive and consolidating knowledge flows in Nobel-winning scientific breakthroughs. *Quantitative Science Studies*, 1–17. https://doi.org/10.1162/qss_a_00323

Yang, A. J., Deng, S., Wang, H., Zhang, Y., & Yang, W. (2023a). Disruptive coefficient and 2-step disruptive coefficient: Novel measures for identifying vital nodes in complex networks. *Journal of Informetrics, 17*(3), Article 101411. https://doi.org/10.1016/j.joi.2023.101411

Yang, A. J., Gong, H., Wang, Y., Zhang, C., & Deng, S. (2024a). Rescaling the disruption index reveals the universality of disruption distributions in science. *Scientometrics, 129*(1), 561–580. https://doi.org/10.1007/s11192-023-04889-x

Yang, A. J., Hu, H., Zhao, Y., Wang, H., & Deng, S. (2023b). From consolidation to disruption: A novel way to measure the impact of scientists and identify laureates. *Information Processing & Management, 60*(5), Article 103420. https://doi.org/10.1016/j.ipm.2023.103420

Yang, A. J., Xu, H., Ding, Y., & Liu, M. (2024b). Unveiling the dynamics of team age structure and its impact on scientific innovation. *Scientometrics*. https://doi.org/10.1007/s11192-024-04987-4

Yang, A. J., Yan, X., Hu, H., Hu, H., Kong, J., & Deng, S. (2024c). Are disruptive papers more likely to impact technology and society? *Journal of the Association for Information Science and Technology*. https://doi.org/10.1002/asi.24947

Zeng, A., Shen, Z. S., Zhou, J. L., Wu, J. S., Fan, Y., Wang, Y. G., & Stanley, H. E. (2017). The science of science: From the perspective of complex systems. *Physics Reports, 714*, 1–73. https://doi.org/10.1016/j.physrep.2017.10.001

Zhang, Y., Wang, Y., Du, H., & Havlin, S. (2024). Delayed citation impact of interdisciplinary research. *Journal of Informetrics, 18*(1), Article 101468. https://doi.org/10.1016/j.joi.2023.101468