Information Processing and Management 59 (2022) 102859

Contents lists available at ScienceDirect

Information Processing and Management

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

Towards transdisciplinary impact of scientific publications: A longitudinal, comprehensive, and large-scale analysis on Microsoft Academic Graph

Yong Huang [a], Wei Lu [a], Jialin Liu [b], Qikai Cheng [a], Yi Bu [b,*]

[a] Information Retrieval and Knowledge Mining Laboratory, School of Information Management, Wuhan University, Wuhan, Hubei 430072, China
[b] Department of Information Management, Peking University, Beijing 100871, China

A B S T R A C T

This paper studies the transdisciplinary impact of scientific publications with a longitudinal, comprehensive, and large-scale analysis on the Microsoft Academic Graph (MAG) dataset. More specifically, this paper aims to understand to what extent publications in discipline A have impact on discipline B. To this end, we propose a novel method to characterize the degree to which a publication impacts another discipline instead of its original discipline. We consider the ratio of the number of citations in a certain discipline and that in the original discipline. We also adopt an OLS regression to identify the equation between the ratio and the affinity of discipline pair and find a clear positive relation. This inspires us to categorize a publication's degree of transdisciplinarity by setting up two thresholds, the top 95% and the bottom 95% confident interval curve (of the fitted line). Publications above the top 95% curve is categorized as transdisciplinary ones, those below the bottom 95% curve as domain-specific ones, and those between the two curve as normal publications. This categorization does not require any pre-defined framework for transdisciplinarity and offers an automatic way of definition by data distribution itself. We find that sociology, mathematics, physics, and chemistry account for a great proportion of transdisciplinary publications that influence other domains, and that medicine, biology, economics, and geology have the greatest proportion of domain-specific publications that show impact in the original discipline. Moreover, we observe a negative relation between the number of citations and the proportion of transdisciplinary publications. A longitudinal analysis presents that the proportion of transdisciplinary publications shows a slightly increase trend for years.

1. Background and motivation

In their *Journal of Documentation* literature review, Sugimoto and Weingart (2015) presented a kaleidoscopic view of discipline and disciplinarity from the perspectives of cognitive, social, communicative, separateness, traditional, institutional, and combinations. This is a very important milestone for scientists to understand different aspects of discipline and disciplinarity. In recent decades, a new direction, namely transdisciplinarity, attracts much attention, partly because of the fact that transdisciplinarity allow researchers to exchange innovative ideas and methods (e.g., Bridle, Vrieling, Cardillo, Araya, & Hinojosa, 2013; Xu, Ding, & Malic, 2015) and that an interconnected world needs new approaches to intellectual inquiry that can challenge common disciplinary and institutional boundaries (Davoudi & Pendlebury, 2010). Zhai, Ding, and Wang (2018) argued that transdisciplinary research and disciplines play an "intermediary" (p. 376) role in knowledge diffusion, a process traced and "footprinted" by citations (Jaffe & Trajtenberg, 1996; Min et al., 2018).

To characterize the scientific impact of and how knowledge diffuses in transdisciplinarity, much extant research has proposed quantitative indicators/measurements to understand transdisciplinary scientific impact, among which Liu and Rousseau (2010)

* Corresponding author.
E-mail address: buyi@pku.edu.cn (Y. Bu).

https://doi.org/10.1016/j.ipm.2021.102859
Received 29 July 2021; Received in revised form 8 December 2021; Accepted 27 December 2021
Available online 3 January 2022

regarded transdisciplinary scientific impact as the final output of transdisciplinary knowledge diffusion. They defined two indicators, namely diffusion breadth and intensity, to quantify the knowledge diffusion among disciplines. Specifically, diffusion breadth is defined as the number of fields where a set of publications are cited, while diffusion intensity is the number of citing publications in a certain field. Regretfully, they did not apply their proposed indicators in a large-scale dataset. Yan, Ding, Cronin, and Leydesdorff (2013) proposed a three-dimensional framework to signify scientific trading, defined as a process containing knowledge input and output. The three dimensions were incoming citation counts (trading impact), the ratio of cited and citing counts (the balance of knowledge "import" and "export"), and self-citation ratios (the degree of self-dependency). Xu, Min, Huang, and Bu (2021) compared the scientific impact of publications generating impact in one and more-than-one domains by investigating a multi-generation citation structure; average clustering coefficient, structural popularity, structural virality, and network density were investigated. They concluded that for papers having a similar "direct" scientific impact, unidisciplinary papers have a relatively greater "indirect" impact, which indicates that the scientific impact of unidisciplinary papers is "deeper" and more persistent, and that compared to unidisciplinary papers, there exists at least a short time in which transdisciplinary papers trigger more follow-up discussions. Zhou, Guns, and Engels (2021) investigated how transdisciplinary knowledge diffusion evolves over time, with a particular interest in its relations to scientific impact and disruptiveness. Diffusion factors for evaluating journals' performance have also been proposed as well (Faber Frandsen, Rousseau, & Rowlands, 2006; Rowlands, 2002). Another thing that worthy noticing is that measurements for transdisciplinary scientific impact and knowledge diffusion are sometimes conducted based on diversity. For example, Liu, Rafols, and Rousseau (2012) recalled the Rao-Stirling to measure how diverse the diffusion of a publication is, whilst Yan (2016) employed the Shannon's entropy to measure discipline-level knowledge diffusion. Other related studies include those of Leydesdorff and Rafols (2011), Liu and Rousseau (2010), and Bu, Li, Gu, and Huang (2021).

There are also many previous empirical studies that present some interesting phenomena in transdisciplinary scientific impact and knowledge diffusion (Hossain, 2020; Liu, Yi, Li, & Li, 2021; Lopez-Olmedo & Gutierrez-Serrano, 2021; Pal, 2020; Xu et al., 2018). Particularly, researchers have been keen to investigate whether knowledge from a certain discipline tends to diffuse to another discipline. Among them, Van Leeuwen and Tijssen (2000) argued that knowledge in a certain discipline tends to diffuse to its adjacent domains. Such observations were specified using empirical data by Leydesdorff and Probst (2009), who concluded that many pieces of knowledge from communication science diffuse to political science and social psychology by examining the journal citation network. Borgman and Rice (1992) observed asymmetric knowledge flows between the fields of library and information science and communication science, in which knowledge from the latter discipline flows to the former but not oppositely. Zhai et al. (2018) investigated the LDA article diffusion and, together with their more recent work (Zhai, Ding, & Zhang, 2021), found a great number of disciplines that have been "affected" by the LDA publication besides computer science itself, such as mathematics, social sciences, and engineering. Similar research also includes that of Cronin and Meho (2008), Levitt, Thelwall, and Oppenheim (2011), and Zhao and Wu (2014).

In network science, epidemic models are a set of means of describing the diffusion of communicable disease or information through individuals in the network, such as susceptible-infected (SI) and susceptible-infected-susceptible (SIS) models (Barabási, 2016). Under the context of transdisciplinary scientific impact and knowledge diffusion, epidemic models have also been employed as a useful tool to understand the dynamic of impact. Kiss, Broom, Craze, and Rafols (2010), for instance, studied the dynamic process of topic diffusion across disciplines through epidemic models. To achieve this, they first defined three statuses of a subject, susceptible (either not aware of a certain research topic or not adopt it, annotated as S), incubating (aware of a certain research topic and "have moved onto actively engaging with it," which "result in tangible research output in the form of papers" [p. 77], annotated as E), and infected (publishing in a certain research topic, annotated as I), and proposed two epidemic models, namely S-E-I and S-I models. As a proof-of-concept study, Kiss et al. (2010) implemented simulations that show a good fit between the simulated and real data. Similarly, Gao and Guan (2012) also implemented an empirical model in their paper to explore how *h*-index-related research had spread. However, one of the problems of these studies adopting epidemic models is that some parameters and thresholds in the models were determined manually without more in-depth discussions.

Going back to transdisciplinary scientific impact, we should realize that its definition varies (Gibbons & Nowotny, 2001; Klein, 2004; Max-Neef, 2005; Nicolescu, 2002, 2014). There are at least two aspects for the concept of transdisciplinarity. On the one hand, the discipline a publication lies in might be transdisciplinary. For instance, a publication in computational social sciences itself combines both social science and quantitative and computational methods. On the other hand, a certain publication might focus on one single discipline; however, it generates impact (later) in other disciplines instead of simply in the original discipline of the publication. For the aforementioned two aspects of transdisciplinarity, this paper pays particular attention on the second aspect (which is often neglected in previous studies), i.e., to understand to what extent publications in discipline A have impact on discipline B. To this end, we propose a novel method to characterize the degree to which a publication impacts another discipline instead of its original discipline. We consider the ratio of the number of citations in a certain discipline and that in the original discipline and aggregate it to the discipline level. We also adopt an Ordinary Least Squares (OLS) regression to identify the equation between the ratio and the affinity of discipline pair and find a clear positive relation. This inspires us to categorize a publication's degree of transdisciplinarity by setting up two thresholds, the top 95% and the bottom 95% confident interval curve (of the fitted line). Publications above the top 95% curve is categorized as transdisciplinary ones, those below the bottom 95% curve as domain-specific ones, and those between the two curve as normal publications. This categorization does not require any pre-defined framework for transdisciplinarity (e.g., the variety-balance-disparity dimensions Leydesdorff & Rafols, 2011; Leydesdorff, Wagner, & Bornmann, 2019; Zhang, Rousseau, & Glänzel, 2016) and offers an automatic way of definition by data distribution itself. We find that sociology, mathematics, physics, and chemistry account for a great proportion of transdisciplinary publications that influence other domains, and that medicine, biology, economics, and geology have the greatest proportion of domain-specific publications that show impact in the original discipline. We observe a negative relation between the number of citations and the proportion of transdisciplinary publications, even with some robustness tests. Moreover, a temporal analysis presents that the proportion of transdisciplinary publications shows a slightly increase trend for years.

2. DATA

The Microsoft Academic Graph (MAG) (Sinha et al., 2015; Wang et al., 2020) is adopted as the empirical dataset of the current paper. To ensure that publications have enough citing records, we select publications published in 1971–2010 with at least one citation, which covers 39,790,276 publications in our empirical study. In the following empirical study, we use all citation records ending in 2016—this selection offers at least five years to accumulate citations.[1] Fig. 1(a) shows the distribution of the number of publications in over years where we particularly annotate 1970, 2010, and 2016 in the sub-figure.

Each of these ~100 million publications are labeled as multiple fields that form a six-level hierarchical structure (i.e., Levels 0, 1, 2, 3, 4, and 5). We follow previous studies (AlShebli, Rahwan, & Woon, 2018; Bu et al., 2021) and select Level 0 (the highest level) as field-of-study labels. This yields 19 macro-level disciplines. Fig. 1(b) presents the number of publications in each discipline. We can see that Medicine and Chemistry occupy the greatest number of publications among all disciplines.

Nevertheless, some of the 19 disciplines defined by MAG seem "transdisciplinary" themselves (i.e., they are not quite clear as a single "discipline") by the fact that publications in these disciplines do not cite other publications in the same discipline quite often. For instance, we notice that publications in Geography have a discipline-level self-citation rate of only 10%. This reminds us that we should remove these disciplines from our further analyses as their degree of "disciplinarity" is not quite high (Zhang et al., 2016). To this end, disciplines with fewer than 30% self-citation rate are eliminated. This results in 12 disciplines out of 19, namely Biology, Chemistry, Computer Science, Economics, Engineering, Geology, Materials Science, Mathematics, Medicine, Physics, Psychology, and Sociology. The mutual citation heat map is shown in Fig. 2.[2] In the figure, excluded disciplines are marked in red rectangles.

3. Methods

Suppose that publication p_{D_0} is labeled as discipline D_0. The scientific impact of p_{D_0} in D_0 could be quantified as the number of citations from other publications in D_0 to p_{D_0}, annotated as:

$$I_{p_{D_0}\to D_0} = C_n\left(p_{D_0}, D_0\right) \tag{1}$$

where n refers to the length of citation window. The cumulative distribution of $C_n(p_{D_0}, D_0)$ can be found in Fig. A1 in the **Appendix**.

Besides, it is likely that p_{D_0} receives citations from other disciplines out of D_0, say D_k ($k = 1, 2, \ldots, 11$ in the scenario of the current empirical study). We characterize the scientific impact of p_{D_0} in D_k with its number of citations from publications in D_k, and annotate as:

$$I_{p_{D_0}\to D_k} = C_n\left(p_{D_0}, D_k\right) \tag{2}$$

We then define the transdisciplinary impact of p_{D_0} (when considering D_k) as the ratio of $I_{p_{D_0}\to D_k}$ and $I_{p_{D_0}\to D_0}$ to indicate to what extent the scientific impact of p_{D_0} in D_0 diffuses to D_k, annotated as:

$$TDI_{p_{D_0}\to D_k} = \frac{I_{p_{D_0}\to D_k}}{I_{p_{D_0}\to D_0}} \tag{3}$$

Note that p_{D_0} may derive multiple $TDI_{p_{D_0}\to D_k}$ when we consider different disciplines (i.e., k). In Eq. (3), empirical studies would remove all publications that have zero citations from D_0 to avoid zero denominators.

$TDI_{p_{D_0}\to D_k}$ implies the scientific impact of p_{D_0} on discipline D_k compared with its original discipline D_0. We can then aggregate $TDI_{p_{D_0}\to D_k}$ of all publications in D_0 as $TDI_{D_0\to D_k}$:

$$TDI_{D_0\to D_k} = \sum_{p_{D_0}\in D_0} TDI_{p_{D_0}\to D_k} \tag{4}$$

Normally, $TDI_{D_0\to D_k}$ relates to the affinity of D_k and D_0. For instance, it might be easier for a physics publication to have impact on biology but more difficult to impact arts, partly because of the field similarity among physics, biology, and arts. This inspires us to link

[1] MAG records all publications in an iterative way. That is to say, at the beginning of 2018, NOT all 2017-or-before publications are fully recorded because of, for instance, publication delay or other non-academic reasons. Such a delay may affect 3–5 years' publications. Thus, a safe strategy is to select the dataset that is 5-or-more years older. This is why we select 2016 as the end of the dataset.

[2] In Fig. 2, the value of each cell is calculated as the percentage of a publication in the ith row citing a publication in the jth column. Thus, the main diagonal value indicates the self-citation rate of a certain discipline. In a few cases, a publication may have multiple Level 0 labels; in these circumstances, we randomly select one Level 0 discipline for this publication to finalize the matrix in Fig. 2. Yet, we should admit that adopting the discipline-level self-citation rate as a signal for removing some disciplines in MAG may cause some biases as MAG does not cover many monographs which is quite important in Arts and Humanities, as well as Social Sciences. This is why the final 12 disciplines are all STEM except Sociology.

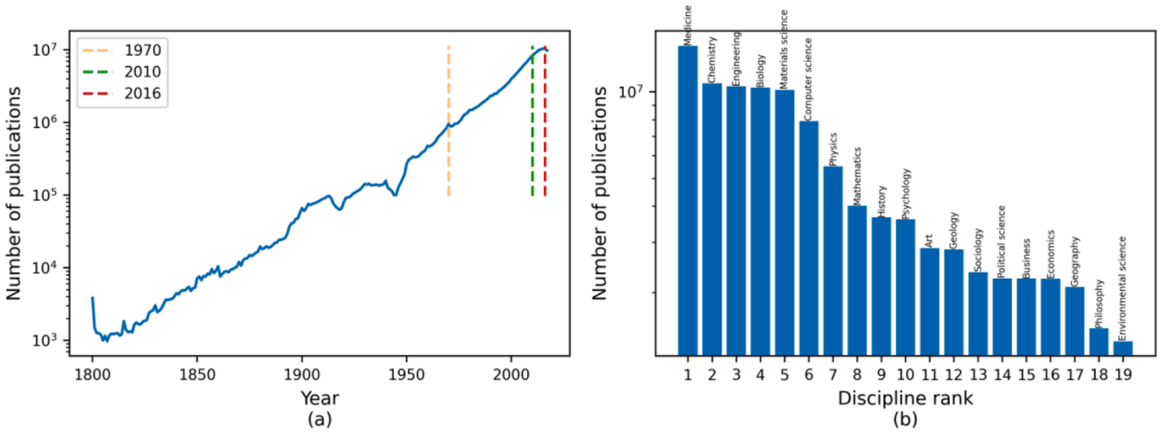

Fig. 1. Distribution of the number of publications in each selected year (a) and each discipline (b).

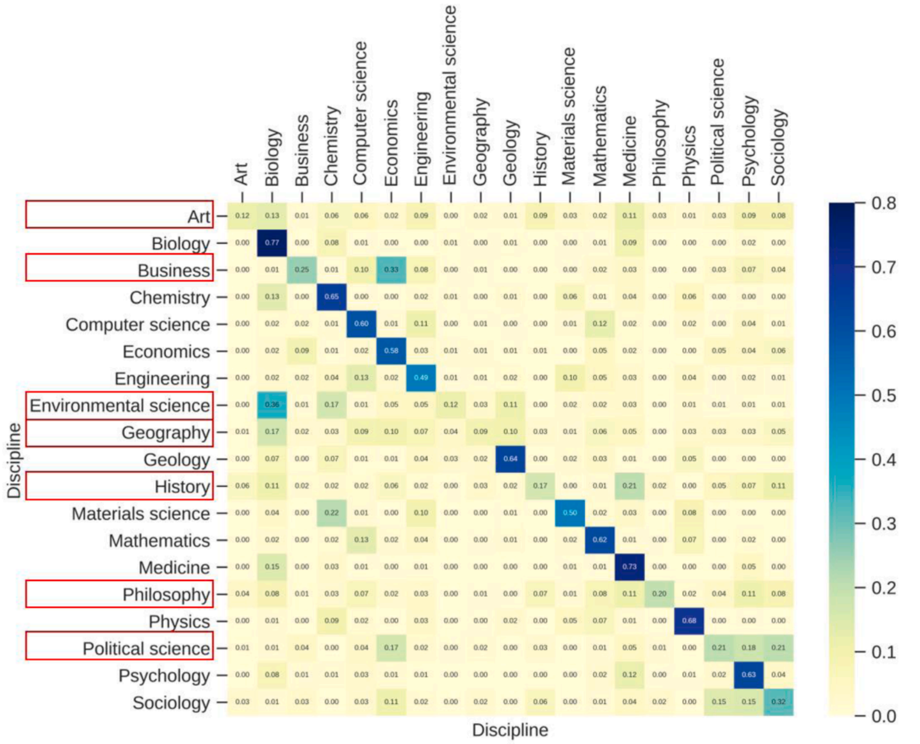

Fig. 2. Mutual citation heat map for all Level 0 disciplines.

$TDI_{D_0 \to D_k}$ with the affinity of D_k on D_0, annotated as $aff(D_0, D_k)$ with a function $F(\cdot)$, i.e.:

$$F(aff(D_0, D_k)) = TDI_{D_0 \to D_k} \tag{5}$$

Here, $aff(D_0, D_k)$ is quantified as:

$$aff(D_0, D_k) = \frac{C_n(D_k \to D_0)}{C_n(D_k \to *)} \tag{6}$$

where $C_n(D_k \to D_0)$ is the number of citations from D_k to D_0 and $C_n(D_k \to *)$ is the total number of references of publications in D_k. From Eq. (6), we see that $aff(D_0, D_k)$ and $aff(D_k, D_0)$ may have quite different values, which shows the affinity of $aff(\cdot)$and echoes many previous scientometric definitions (e.g., Chinchilla-Rodríguez, Bu, Robinson-García, & Sugimoto, 2021; Yan et al., 2013).

Note that in Eq. (6), the denominator considers to what extent D_k endorses all disciplines (a global structure) and the nominator to what extent D_k endorses D_0.

We operationalize the above definitions on MAG and plot the relation between transdisciplinary impact ($TDI_{D_0 \to D_k}$, the vertical axis) and discipline affinity ($aff(D_0, D_k)$, the horizontal axis) in a two-dimensional map in Fig. 3. Here, each data point represents a pair of disciplines. We adopt an OLS regression model to fit a straight line (in blue) in double logarithmic scale and identify the top and bottom 95% confident interval lines (in red straight lines) as well in Fig. 3. For example, we see in Fig. 3 that the data point representing Medicine and Biology posits exactly adjacent to the fitted straight curve. The discipline affinity of the two selected disciplines is quite high, which make sense because many medical publications have their own biological principles (foundations). We also observe that the data point representing Physics and Psychology lies outside the top 95% confident interval curve, indicating that this discipline pair generates quite a transdisciplinary impact. Although the discipline affinity is not quite high for the discipline pair, their $TDI_{D_0 \to D_k}$ seems to outperform normal discipline pairs' than expected.

Fig. 3 presents that $\log(TDI_{D_0 \to D_k}) = 0.0661 * \log(aff(D_0, D_k)) - 0.1783$ with $R^2 = 0.064$.[3] With this fitted curve, we category a certain publication based upon its transdisciplinary impact into three groups. Specifically, if any of its $TDI_{p_{D_0} \to D_k}$ is greater than the top 95% confident interval threshold, this paper is classified as an "**transdisciplinary**" publication; if all of its $TDI_{p_{D_0} \to D_k}$ is smaller than the bottom 95% confident interval threshold, this paper is classified as a "**domain-specific**" publication; and if any of these two criteria does not match, this paper is classified as a "**normal**" publication.

The determination of 95% confident interval could be revised based upon various scenarios of future studies. 95% is selected in the current study only because this is a commonly used threshold. The three-fold classification may change to, for example, five folds, if two distinct thresholds are set. Yet, one should carefully determine the choice of these parameters and thresholds as they may cause the paradox of interpretations on empirical results.

4. Empirical results

Fig. 4 shows the proportion of transdisciplinary (in green), normal (in yellow), and domain-specific (in red) publications in different disciplines, where we can see that in all disciplines, domain-specific publications occupy for the most. We also see that sociology features with the greatest proportion of transdisciplinary publications (12.77%). Indeed, the increasing transdisciplinarity of sociology has been explicitly pointed out by many sociologists such as Letherby (2005), together discussed with the fragmentation characteristics of this discipline (O'Reilly, 2009). Besides sociology, mathematics (12.11%), physics (11.26%), and chemistry (9.93%) also account for a great proportion of transdisciplinary publications, indicating that publications in these disciplines have more tendencies to diffuse to other disciplines. Take the mathematics as an example: As concluded by Frank, Wang, Cebrian, and Rahwan (2019), mathematics are increasingly referenced by researchers in, for example, artificial intelligence. Regarding domain-specific publications, we observe that medicine (74.44%), biology (71.62%), economics (70.56%), and geology (70.25%) have the greatest proportion of unidisciplinary publications that show impact in the original discipline. Moreover, in general, when not considering discipline-level differences, the proportions of transdisciplinary, normal, and domain-specific publications are 7.56%, 25.67%, and 66.78%, respectively. These aggregated results imply that overall the majority of publications is quite domain-specific and that transdisciplinary publications are still quite limited in general.

Fig. 5 shows the proportion of three types of publications with different numbers of citations. We group publications using its citation count (within a five-year-long citation time window) as [1,5], (5,10], (10,20], (20,50], (50,100], and more than 100. We clearly see that publications with a greater number of citations tend to be domain-specific ones: Specifically, there are only ~63.18% domain-specific publications with one to five citations, but that number jumps to 86.51% for publications with 100+ citations. This finding hints the deprivation of transdisciplinary publications regarding scientific impact because we do observe a smaller proportion of transdisciplinary publications in highly cited groups, which is consistent with what Levitt and Thelwall (2008) found in their empirical study covering life sciences, health sciences, and physical sciences. On the other hand, such a finding implies that most of the citations of highly cited publications come from their original domains and that disciplinary boundaries may still be there (Klein, 1996). This is reasonable because transdisciplinary knowledge diffusion needs additional time and costs, such as understanding jargons in distinct disciplines (Xu et al., 2015).

Yet, the current finding does not consider the interaction between discipline and citation count—for instance, biology publications tend to be cited more than mathematic—nor the discipline-wise publication size, and thus may bias some highly productive and cited disciplines. To this end, we normalize the number of citations of a certain publication by considering its discipline and derive the citation ranking percentile of this publication. Note that similar to aforementioned empirical study, here we take into consideration a five-year-long citation time window to keep the consistency. The updated results are shown in Fig. 6, where 0%−20% citation rank indicates higher cited publications (in their disciplines), while 80%+ shows lower cited ones. We see that the negative relation between citation impact and the proportion of transdisciplinary publications still exists. Particularly, there are 58.88% publications that are domain-specific for publications ranked top 20% in their own disciplines; yet, such percentage equals 61.18% for those ranking between 40%−60%. As for quite lowly cited publications (those ranked the most bottom 20% publications), the proportion of domain-

[3] Indeed, as shown in Fig. 3, the value of R^2 is not quite great because we fit the (blue) straight line using EACH DATA POINT in the figure. We also attempted the average value (shown as a red dotted curve) by aggregating $TDI_{D_0 \to D_k}$ with data points within the same bin of discipline affinity and obtained an R^2 of 0.264. However, we still adopted the former strategy, i.e., fitting the curve using EACH DATA POINT, in that we are hoping to know the upper and the lower boundaries of $TDI_{D_0 \to D_k}$ with the 95% threshold, which further derives the Point Percentage in Range within the 95% threshold. We found that 96.2% data points were included within the two thresholds.

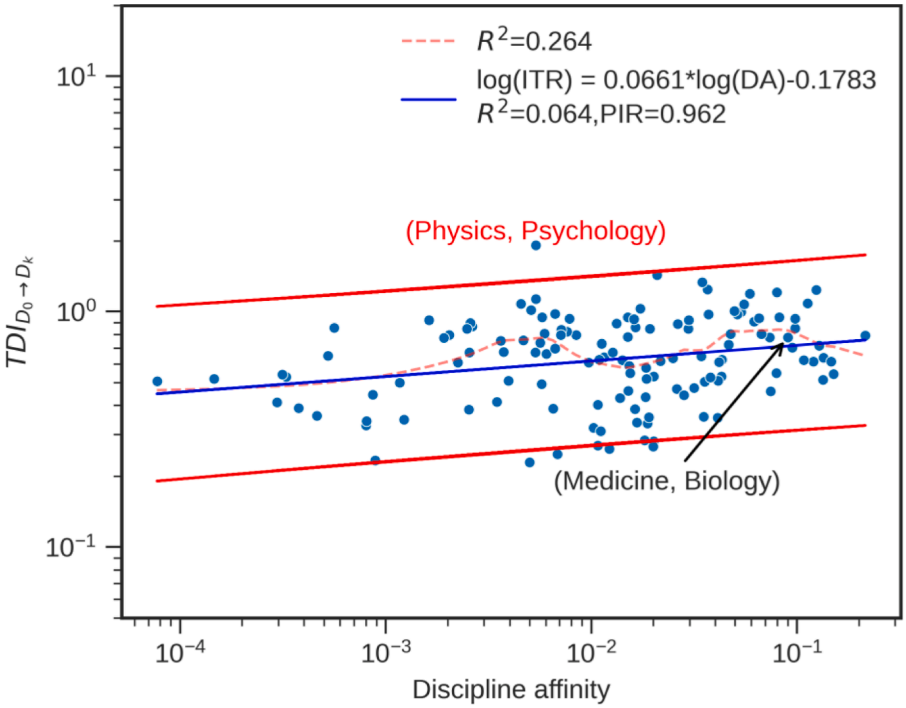

Fig. 3. Relation between transdisciplinary impact ($TDI_{D_0 \to D_k}$) and discipline affinity ($aff(D_0, D_k)$). Each data point represents a pair of disciplines. The *p*-value for the hypothesis test of the regression analysis equals 0.004, indicating a statistical significance.

specific publications is quite great (77.49%). The only exception is 40%–60% citation ranking, which have a greater proportion of transdisciplinary publications than the 20%–40% group. From these observations, domain-specific publications have a weaker advantage in terms of their citation impact after normalization.

We further examine how the proportion of transdisciplinary, normal, and domain-specific publications evolve over years. As shown in Fig. 7, the proportion of transdisciplinary publications shows a slightly increase trend for years, but such changes are quite nuanced—from ~0.05 in 1970 to ~0.08 in 2010, which shows that in the past 50 years fewer than one out of ten publications are quite transdisciplinary. The proportions of normal and domain-specific publications increase and decrease, respectively, indicating that there are many more publications that generate transdisciplinary impact recently. Yet, one should note that such changes are not quite obvious. This finding reminds us that although transdisciplinarity has been encouraged for years, new knowledge and ideas still come from a unidisciplinary perspective.

Furthermore, the proportion (%) of transdisciplinary, normal, and domain-specific publications over years for different disciplines is shown in Fig. A2 in the **Appendix**.

As a comparison of the proposed measurement, we adopted the DIV indicator proposed by Leydesdorff et al. (2019) in their *Journal of Informetrics* paper; this indicator combines three dimensions of diversities (i.e., variety, balance, and disparity) together into one single measurement. For each paper, we calculate DIV for each publication using their citing publications (the so-called "cited" side mentioned in Leydesdorff et al. (2019)), as well as its maximum $TDI_{p_{D_0} \to D_k}$. As the two variables do not follow a normal distribution, we further quantify the Spearman's correlation coefficient between them. The coefficient equals -0.38 (p-value $= 0.00$).[4] The value of the Spearman's correlation coefficient seems low (and negative), which implies that ours offers a new dimension for understanding transdisciplinarity.

5. Discussion

Transdisciplinarity has been widely and strongly encouraged for many years in almost all major countries. Klein and Falk-Krzensinski (2017) mentioned that the Computing Research Association encourages young scholars in Information Science, Computing, and Engineering to highlight transdisciplinarity in their job interviews and to highlight their proposed collaboration-based center/-institute to "seek advice on how to balance participation on large team projects with work that establishes a strong individual reputation" (p. 1056). However, we, surprisingly, find that highly cited publications tend to have a fewer probability of transdisciplinary citations. While in the tide of transdisciplinarity, this finding inspires future science funding providers that they are not

[4] We should also mention that ours highlights the maximum and existing measurements mostly adopt a balance-based strategy.

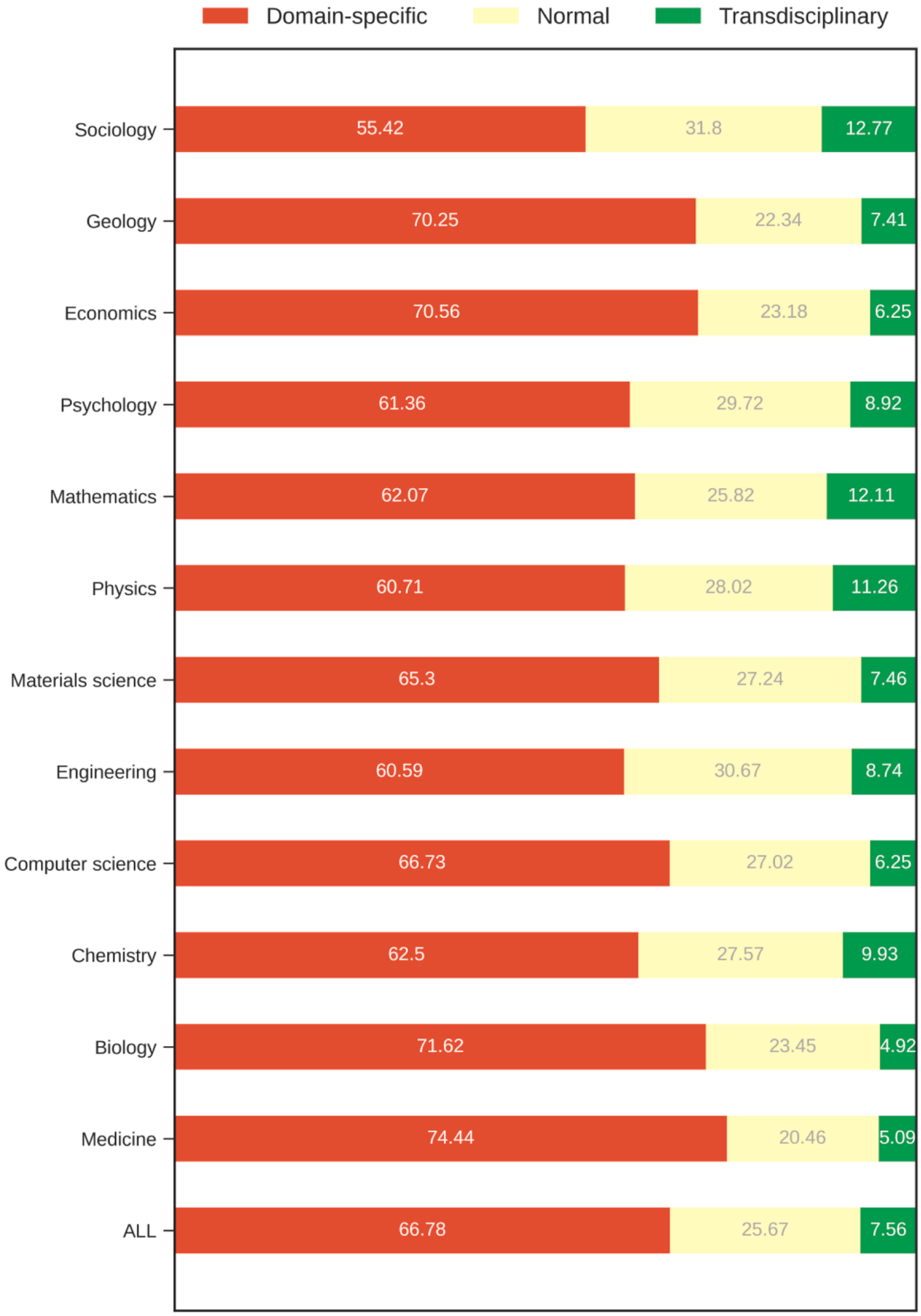

Fig. 4. Proportion (%) of transdisciplinary, normal, and domain-specific publications in different disciplines. The red, yellow, and green bar represent the proportions of domain-specific, normal, and transdisciplinary publications.

ought to neglect financial support of unidisciplinary studies. After we normalize the citation count by their disciplines (i.e., citation ranking), we still see a consistent negative relation, indicating the robustness of the finding.

We also see that different disciplines have various proportions of transdisciplinary scientific impact. We observe that sociology, mathematics, physics, and chemistry feature with the greatest proportion of publications with transdisciplinary impact. Pedagogically, training programs in these disciplines should include more transdisciplinary coursework and/or transdisciplinary research practices to educate trainees to understand the jargons, writing styles, and research contexts of its adjunct disciplines; this is to guarantee that their future studies could be easily understood by and diffused to other disciplines and that the scientific impact of their publications stride

Fig. 5. Proportion of three types of publications with different numbers of citations. The red, yellow, and green bar represent the proportions of domain-specific, normal, and transdisciplinary publications. Publications are grouped based upon their numbers of citations within five years after published.

across disciplinary boundaries.

Compared with existing measurements for transdisciplinarity, ours has several distinctions.

1 A dynamic perspective: Most existing measurements employ references of a focal publication as a proxy of its transdisciplinarity, which is static—the degree of transdisciplinarity of a certain publication is fixed once published. Nonetheless, the current one, similar to the DIV-cited indicator first proposed by Leydesdorff et al. (2019), offers a cited angle and paints a more dynamic and

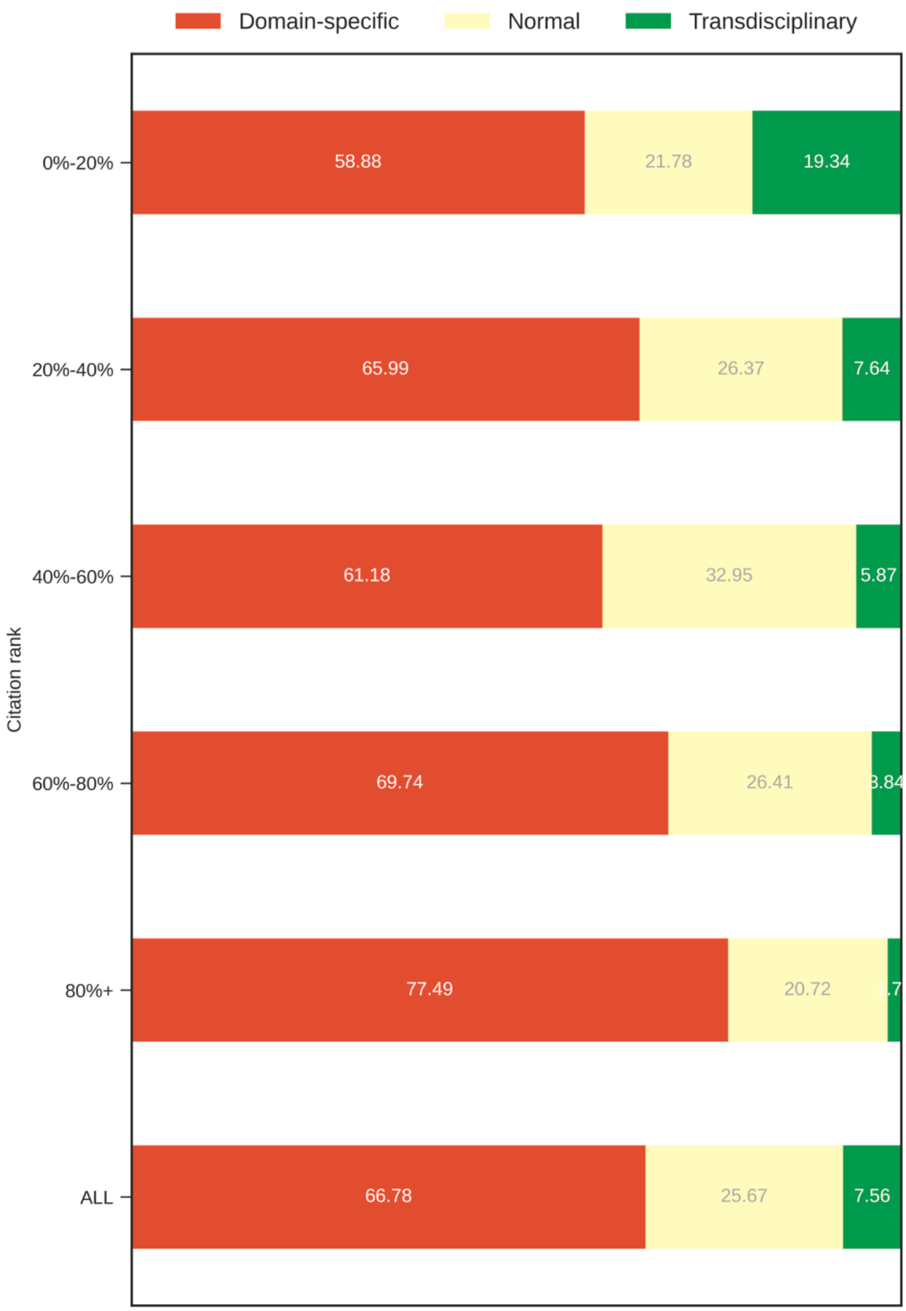

Fig. 6. Proportion of three types of publications with different citation ranks. The citation rank is calculated for publications in one single discipline. Citations are quantified in a five-year-long time window. The red, yellow, and green bar represent the proportions of domain-specific, normal, and transdisciplinary publications. 0%−20% citation rank indicates higher cited publications (in their disciplines), while 80%+ shows lower cited ones.

nuanced picture on how the degree of transdisiciplinarity and transdisciplinary scientific impact evolve over time. Also, the dynamic perspective is also revealed in terms of the perspective of *discipline pairs* instead of *disciplines*.

2 A global structure: Although extant measurements, such as DIV (Leydesdorff et al., 2019), consider disparity as a factor, only disciplines being cited and/or citing a focal publication are taken into consideration. Yet, the current measurement involves all discipline pairs with a global disciplinary picture, which may offer a more accurate and comprehensive view for understanding transdisciplinarity. Moreover, the degree of transdisciplinarity, in this paper, is derived from a global pattern by fitting an OLS regression curve with two 95% confident intervals. Such a global-to-local calculation ensures that any global-level changes would influence the results of individual-level transdisciplinarity (Börner et al., 2012).

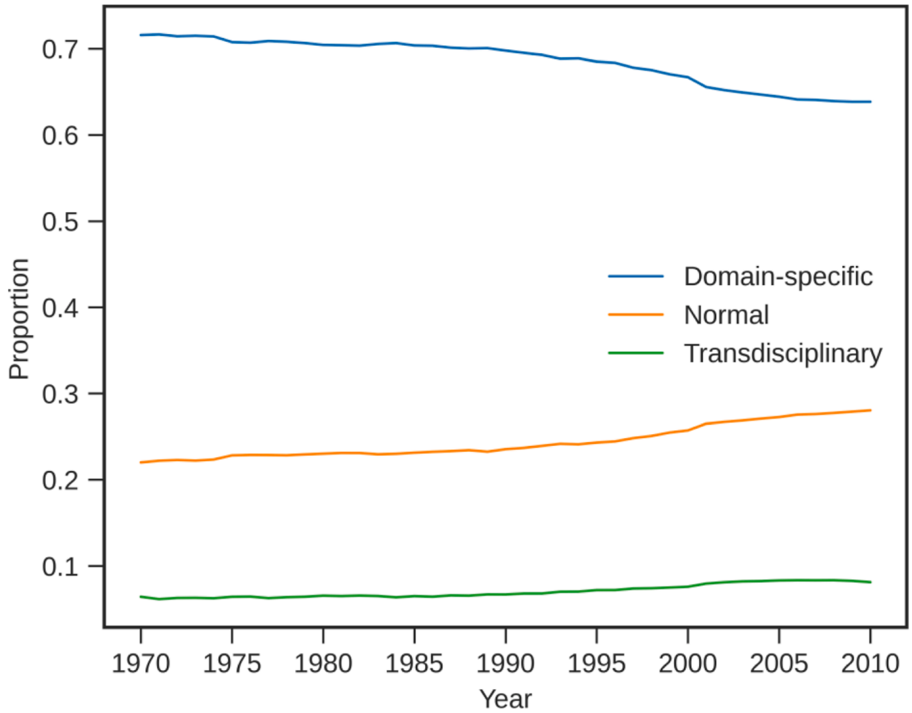

Fig. 7. Proportion (%) of transdisciplinary, normal, and domain-specific publications over years.

3 Clear thresholds: The proposed measurement explicitly determines a threshold that distinguishes "transdisciplinary" and "normal" publications. A clear, yet adjustable, threshold assists science policy makers and funding providers to identify which stakeholders, considering their academic records, should belong to the "transdisciplinary" group when, for instance, doing decision-making or considering funding assignment.

4 Asymmetry of disciplines: Many existing indicators on transdisciplinarity have proposed three different dimensions, namely variety (number of categories), balance (evenness of categories), and disparity (dissimilarity among categories), to quantify transdisciplinarity (essentially topical/disciplinary diversity). When considering this framework, this paper supplements the dimension of "disparity" in that we consider its asymmetry. Like what Zhang et al. (2016) have done, we quantify the strength of discipline-level citation/reference flows as a proxy of characterizing the affinity/similarity between disciplines. Yet, the major distinction is that our defined affinity is asymmetric while Zhang et al. (2016) symmetric. The asymmetry reveals the relative strength of citation linkages in science, which allows for the calibration of relative importance between disciplines, measuring the amount of citing relationships (Chinchilla-Rodríguez et al., 2021) as well as the strength of transdisciplinary knowledge diffusion.

6. Conclusions

This paper proposes a new direction of quantifying the degree of transdisciplinarity of a certain publication from the perspective of transdisciplinary impact. We see one set of disciplines (e.g., sociology, mathematics, physics, and chemistry) with a greater proportion of transdisciplinary publications that influence other discipline and another set of disciplines (e.g., medicine, biology, economics, and geology) with a lower proportion. Besides, citation impact and the proportion of transdisciplinarity have negative relations, and a temporal analysis shows that the proportion of transdisciplinary publications shows a slightly increase trend for years.

We observe a low correlation between DIV and ours. The main reason why ours offers a different dimension from existing indicators, such as DIV, is that they emphasize a diversity perspective focusing on how "wide" the impact is. Yet, our indicator calculates maximum values regarding impacts among disciplines, highlighting how "strong" the impact is in another discipline. That's why, in our empirical explorations, we do find that many top ranked papers under DIV come from multi-disciplinary journals, but in ours, we observe many papers from mono-disciplinary journals (but these papers do generate quite strong transdisciplinary impact).

There are several limitations in the current study. For example, we remove Level 0 disciplines in MAG if a discipline has fewer than 30% discipline-level self-citations. This leads to the inexistence of Arts, Humanities, and (partial) Social Sciences in our empirical study except Sociology. We should admit that these disciplines contain many monographs, regional or national serials, and/or reports (Broadus, 1971) and may have distinct referencing behavior compared with STEM (Lariviere & Sugimoto, 2019). Future studies should separately consider Arts, Humanities, and Social Sciences and examine potential differences regarding their transdisciplinary scientific impact from STEM's, especially considering document type disparity.

Meanwhile, the selection of 30% as a threshold is arbitrary, purely based on a manual check of the disciplinary self-citation distribution. Furthermore, the current empirical study does not set up dynamic citation time windows for different disciplines, which may cause some biases, especially those with a longer half-life. The configuration of citation time window should be carefully established with sophisticated analyses on discipline-wise obsolescence curve (Gou, Meng, Chinchilla-Rodríguez, & Bu, 2021).

CRediT authorship contribution statement

Yong Huang: Conceptualization, Writing – original draft, Data curation, Investigation. **Wei Lu:** Conceptualization, Supervision, Investigation. **Jialin Liu:** Writing – review & editing, Investigation. **Qikai Cheng:** Writing – review & editing, Visualization. **Yi Bu:** Conceptualization, Writing – original draft, Writing – review & editing.

Acknowledgments

This work was supported by High-performance Computing Platform of Peking University and was financially supported by the Youth Program of National Natural Science Foundation of China (#72004168 and #72104007) and the Youth Project of Humanities and Social Sciences of the Ministry of Education (MOE) of China (#21YJC870001). The authors would like to thank the two anonymous reviewers and all members of the Knowledge Discovery Laboratory (KD Lab) at Peking University for their insightful suggestions.

Appendix. None

The appendix contains two figures as shown below.

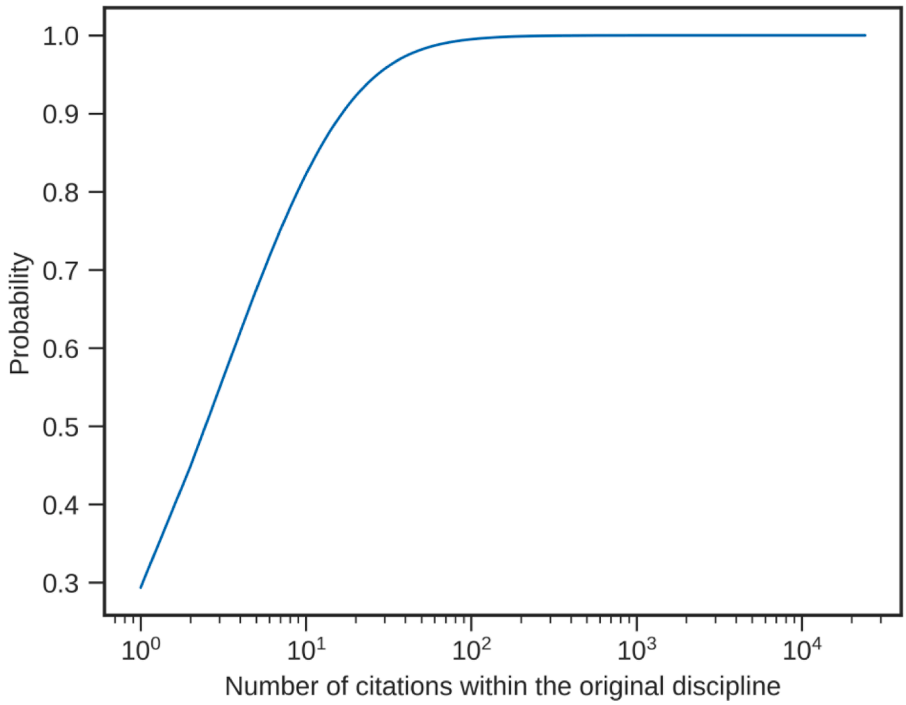

Fig. A1. Cumulative distribution function (CDF) of the number of citations within the original discipline using a five-year-long time window.

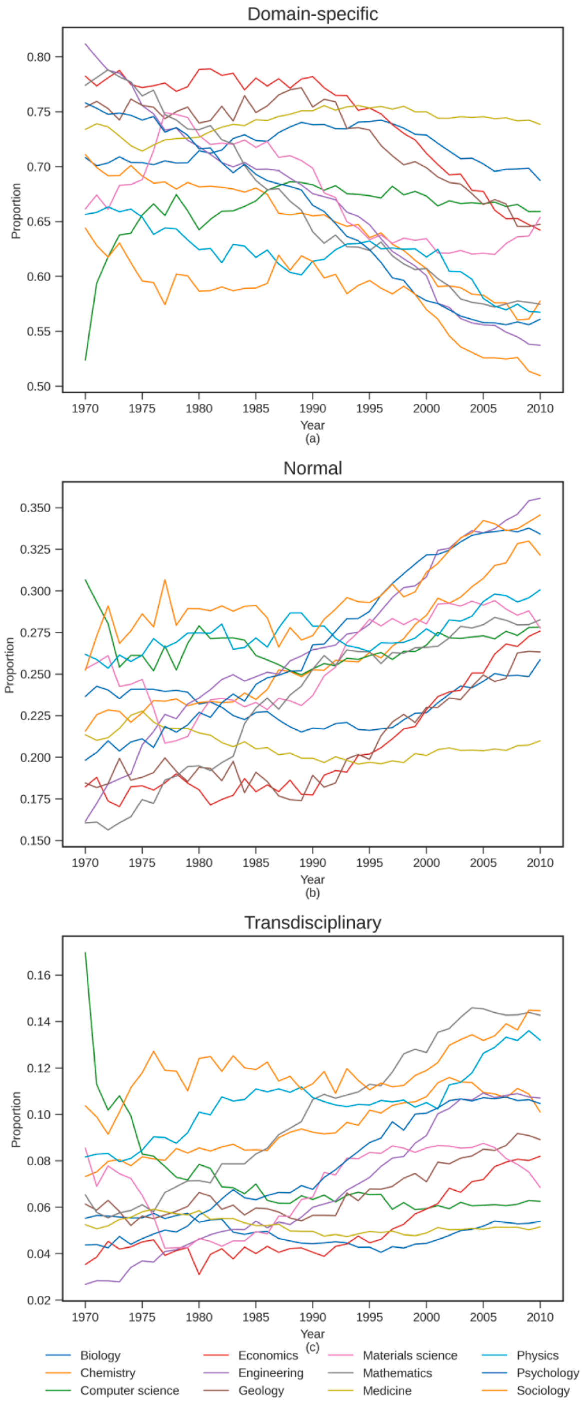

Fig. A2. Proportion (%) of transdisciplinary (a), normal (b), and domain-specific (c) publications over years for different disciplines.

References

AlShebli, B. K., Rahwan, T., & Woon, W. L. (2018). The preeminence of ethnic diversity in scientific collaboration. *Nature Communications, 9*(1), 1–10. https://doi.org/10.1038/s41467-018-07634-8

Barabási, A. L. (2016). *Network science*. Cambridge University Press.

Borgman, C. L., & Rice, R. E. (1992). The convergence of information science and communication: A bibliometric analysis. *Journal of the American Society for Information Science, 43*(6), 397–411.

Börner, K., Klavans, R., Patek, M., Zoss, A. M., Biberstine, J. R., Light, R. P., et al. (2012). Design and update of a classification system: The UCSD map of science. *PloS One, 7*(7).

Bridle, H., Vrieling, A., Cardillo, M., Araya, Y., & Hinojosa, L. (2013). Preparing for an interdisciplinary future: A perspective from early-career researchers. *Futures, 53*, 22–32.

Broadus, R. N. (1971). Literature of social sciences-survey of citation studies. *International Social Science Journal, 23*(2), 236–243.

Bu, Y., Li, M., Gu, W., & Huang, W. (2021). Topic diversity: A discipline scheme-free diversity measurement for journals. *Journal of the Association for Information Science and Technology, 72*(5), 523–539.

Chinchilla-Rodríguez, Z., Bu, Y., Robinson-García, N., & Sugimoto, C. R. (2021). An empirical review of the different variants of the probabilistic affinity index as applied to scientific collaboration. *Scientometrics, 126*(2), 1775–1795.

Cronin, B., & Meho, L. I. (2008). The shifting balance of intellectual trade in information studies. *Journal of the American Society for Information Science and Technology, 59*(4), 551–564.

Davoudi, S., & Pendlebury, J. (2010). Centenary paper: The evolution of planning as an academic discipline. *Town Planning Review, 81*(6), 613–646.

Faber Frandsen, T., Rousseau, R., & Rowlands, I. (2006). Diffusion factors. *Journal of Documentation, 62*(1), 58–72.

Frank, M. R., Wang, D., Cebrian, M., & Rahwan, I. (2019). The evolution of citation graphs in artificial intelligence research. *Nature Machine Intelligence, 1*(2), 79–85.

Gao, X., & Guan, J. (2012). Network model of knowledge diffusion. *Scientometrics, 90*(3), 749–762.

Gibbons, M., & Nowotny, H. (2001). The potential of transdisciplinarity. *Transdisciplinarity: Joint problem solving among science, technology, and society* (pp. 67–80). Springer.

Gou, Z., Meng, F., Chinchilla-Rodríguez, Z., & Bu, Y. (2021, July). Revisiting the obsolescence process of individual scientific publications: Operationalisation and a preliminary cross-discipline exploration. In *Proceedings of the 18th international conference on scientometrics and informetrics (ISSI 2021)*.

Hossain, M.M. (2020) Current status of global research on novel coronavirus disease (COVID-19): A bibliometric analysis and knowledge mapping. Available at SSRN 3547824.

Jaffe, A. B., & Trajtenberg, M. (1996). Flows of knowledge from universities and federal laboratories: Modeling the flow of patent citations over time and across institutional and geographic boundaries. *Proceedings of the National Academy of Sciences, 93*(23), 12671–12677.

Kiss, I. Z., Broom, M., Craze, P. G., & Rafols, I. (2010). Can epidemic models describe the diffusion of topics across disciplines? *Journal of Informetrics, 4*(1), 74–82.

Klein, J. T. (1996). *Crossing boundaries: Knowledge, disciplinarities, and interdisciplinarities*. University of Virginia Press.

Klein, J. T. (2004). Prospects for transdisciplinarity. *Futures, 36*(4), 515–526.

Klein, J. T., & Falk-Krzesinski, H. J. (2017). Interdisciplinary and collaborative work: Framing promotion and tenure practices and policies. *Research Policy, 46*(6), 1055–1061.

Lariviere, V., & Sugimoto, C. R. (2019). The journal impact factor: A brief history, critique, and discussion of adverse effects. *Springer handbook of science and technology indicators* (pp. 3–24). Springer.

Letherby, G. (2005). Current issues and future trends in sociology: Extending the debate in sociological research online. *Sociological Research Online, 10*(1), 69–70.

Levitt, J. M., & Thelwall, M. (2008). Is multidisciplinary research more highly cited? A macrolevel study. *Journal of the American Society for Information Science and Technology, 59*(12), 1973–1984. https://doi.org/10.1002/asi.20914

Levitt, J. M., Thelwall, M., & Oppenheim, C. (2011). Variations between subjects in the extent to which the social sciences have become more interdisciplinary. *Journal of the American Society for Information Science and Technology, 62*(6), 1118–1129. https://doi.org/10.1002/asi.21539

Leydesdorff, L., & Probst, C. (2009). The delineation of an interdisciplinary specialty in terms of a journal set: The case of communication studies. *Journal of the American Society for Information Science and Technology, 60*(8), 1709–1718.

Leydesdorff, L., & Rafols, I. (2011). Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Informetrics, 5*(1), 87–100.

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.

Liu, X., Yi, B., Li, M., & Li, J. (2021). Is interdisciplinary collaboration research more disruptive than monodisciplinary research? *Proceedings of the Association for Information Science and Technology, 58*(1), 264–272.

Liu, Y., Rafols, I., & Rousseau, R. (2012). A framework for knowledge integration and diffusion. *Journal of Documentation, 68*(1), 31–44.

Liu, Y., & Rousseau, R. (2010). Knowledge diffusion through publications and citations: A case study using ESI-fields as unit of diffusion. *Journal of the American Society for Information Science and Technology, 61*(2), 340–351.

Lopez-Olmedo, R., & Gutierrez-Serrano, N. G. (2021). Transdisciplinary knowledge production in mainstream journals: From the perspective of the participation of social actors in Mexico. *Scientometrics, 126*(3), 2627–2641.

Max-Neef, M. A. (2005). Foundations of transdisciplinarity. *Ecological Economics, 53*(1), 5–16.

Min, C., Ding, Y., Li, J., Bu, Y., Pei, L., & Sun, J. (2018). Innovation or imitation: The diffusion of citations. *Journal of the Association for Information Science and Technology, 69*(10), 1271–1282.

Nicolescu, B. (2002). *Manifesto of transdisciplinarity*. Suny Press.

Nicolescu, B. (2014). Methodology of transdisciplinarity. *World futures, 70*(3–4), 186–199.

O'Reilly, K. (2009). For interdisciplinarity and a disciplined, professional sociology. *Innovation: The European Journal of Social Science Research, 22*(2), 219–232. https://doi.org/10.1080/13511610903075761

Pal, J. K. (2020). Understanding research productivity in the realm of evaluative scientometrics. *Annals of Library and Information Studies (ALIS), 67*(1), 67–69.

Rowlands, I. (2002). Journal diffusion factors: A new approach to measuring research influence. *Aslib Proceedings, 54*, 77–84.

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B. P., et al. (2015). An overview of Microsoft Academic Service (MAS) and applications. In *Proceedings of the 24th international conference on world wide web* (pp. 243–246).

Sugimoto, C. R., & Weingart, S. (2015). The kaleidoscope of disciplinarity. *Journal of Documentation*.

Van Leeuwen, T., & Tijssen, R. (2000). Interdisciplinary dynamics of modern science: Analysis of cross-disciplinary citation flows. *Research Evaluation, 9*(3), 183–187.

Wang, K., Shen, Z., Huang, C. Y., Wu, C. H., Dong, Y., & Kanakia, A. (2020). Microsoft Academic Graph: When experts are not enough. *Quantitative Science Studies*, 1–18.

Xu, J., Bu, Y., Ding, Y., Yang, S., Zhang, H., Yu, C., et al. (2018). Understanding the formation of interdisciplinary research from the perspective of keyword evolution: A case study on joint attention. *Scientometrics, 117*(2), 973–995.

Xu, J., Ding, Y., & Malic, V. (2015). Author credit for transdisciplinary collaboration. *PloS One, 10*(9), Article e0137968.

Xu, J., Min, C., Huang, W., & Bu, Y. (2021, July 12). Interdisciplinarity vs. Unidisciplinarity: A structural comparison of multi-generation citations and references. In *Proceedings of the 18th international conference on scientometrics and informetrics (ISSI 2021)*.

Yan, E. (2016). Disciplinary knowledge production and diffusion in science. *Journal of the Association for Information Science and Technology, 67*(9), 2223–2245.

Yan, E., Ding, Y., Cronin, B., & Leydesdorff, L. (2013). A bird's-eye view of scientific trading: Dependency relations among fields of science. *Journal of Informetrics, 7*(2), 249–264.

Zhai, Y., Ding, Y., & Wang, F. (2018). Measuring the diffusion of an innovation: A citation analysis. *Journal of the Association for Information Science and Technology, 69*(3), 368–379.

Zhai, Y., Ding, Y., & Zhang, H. (2021). Innovation adoption: Broadcasting versus virality. *Journal of the Association for Information Science and Technology, 72*(4), 403–416.

Zhang, L., Rousseau, R., & Glänzel, W. (2016). Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account. *Journal of the Association for Information Science and Technology, 67*(5), 1257–1265.

Zhao, R., & Wu, S. (2014). The network pattern of journal knowledge transfer in library and information science in China. *KO Knowledge Organization, 41*(4), 276–287.

Zhou, H., Guns, R., & Engels, T. (2021, July). The evolution of interdisciplinarity in five social sciences and humanities disciplines: Relations to impact and disruptiveness. In *Proceedings of the 18th international conference on scientometrics and informetrics (ISSI 2021)*.